(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 December 2000 (21.12.2000)

PCT

(10) International Publication Number
WO 00/77665 A2

(51) International Patent Classification[7]: G06F 17/00

(21) International Application Number: PCT/US00/16032

(22) International Filing Date: 9 June 2000 (09.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/330,920 11 June 1999 (11.06.1999) US

(71) Applicant: CLINICAL DYNAMICS, INC. [US/US]; 303 Potrero Street, Suite 40-D, Santa Cruz, CA 95060 (US).

(72) Inventor: O'KEEFE, Kelly, Ray; 836 Jewell Avenue, Pacific Grove, CA 93950 (US).

(74) Agents: JAKOPIN, David, A. et al.; Pillsbury Madison & Sutro LLP, 1100 New York Avenue, N.W., Washington, DC 20005 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD AND APPARATUS FOR AUTOMATICALLY ALLOCATING STAFFING

(57) Abstract: The present invention provides a method using a computer to estimate staffing required for a present patient population. The computer operates upon data related to various characteristics regarding the present patient population, such as age, sex, and physical ailment, along with data related to a representative past patient population having similar characteristics and resources allocated to that past patient population, including staffing levels, that were required to care for the past patient population. Using the past patient characteristics and associated allocated resources as a guide, models are created that describe the correlation between that past patient population and the allocated resources. Thereafter, the models are used to determine the staffing levels required to adequately care for the present patient population.

WO 00/77665 A2

METHOD AND APPARATUS FOR AUTOMATICALLY ALLOCATING STAFFING

Field of the Invention

The present invention relates to a method and apparatus for automatically allocating staffing and, more particularly, a method and apparatus for automatically allocating staffing in a health care environment. The method and apparatus is also used for automatically severity adjusting workload data and, more particularly, is a method and apparatus for automatically severity adjusting workload data from a health care environment.

Background of the Invention

A continuing challenge in modern business is to manage and to use resources efficiently. One of the resources that need to be managed is staffing. At all different levels of staffing, such as doctors, nurses, and technicians in the health care environment, if there is too many staff, more resources are expended than necessary. If, however, too few staff are present, not all the work that is required can be completed in time. Thus, in a health care setting, if there is insufficient staff, patient care can be compromised and quality suffers. However, if there are too many staff, resources are not being efficiently used. This is complicated in the healthcare setting by the fact that the workload required to care for a patient is dependent upon how severely ill the patient is.

Still further, in many business environments, the work changes. In the health care environment, for instance, a sudden influx of patients will oftentimes dramatically change the medical staffing needs. These work changes make it difficult to determine optimum staffing at any moment in time.

A traditional approach to this problem is to determine staffing levels using a tiered process. A baseline is established for staffing, usually by a full time equivalent ("FTE") allocation for the work under consideration. This staffing level is refined by having a knowledgeable supervisor estimate the staffing required to meet the work needs for the area in which the supervisor is responsible and then prospectively make minor changes to scheduled staffing levels. Final adjustments are made in real time, by contemporaneously calling additional staff in to manage unanticipated demand or by calling off staff if the work was overestimated. Refinement of the baseline and final adjustments are notable for their subjective nature. The staffing decisions depend upon both the experience of the decision-maker and upon their biases.

What is needed, therefore, is an apparatus and method of more efficiently and objectively allocating staffing resources. What is further needed is an apparatus and method for severity adjusting patient populations so the workload that was actually used to care for them can be compared.

Summary of the Invention

It is an object of the present invention to provide an apparatus and method of more efficiently allocating staffing resources.

It is a further object of the present invention to use data regarding past workload statistics in estimating staffing required for work presently being performed.

It is another object of the present invention to use a computer to estimate the staffing required.

It is an object of the present invention to provide an apparatus and method of more efficiently allocating staffing resources in a health care environment with a changing patient population.

It is a further object of the present invention to use data regarding past patient characteristics and workload statistics in estimating staffing required for a present patient workload.

It is another object of the present invention to use a computer to estimate the staffing required for a present patient population.

It is another object of the present invention to provide an apparatus and method of comparing workload used to care for different patient populations.

It is a further object of the present invention to use data regarding past workload statistics and reference workload data to determine the productivity of healthcare delivery.

It is another object of the present invention to use a computer determine the severity adjusted workload and productivity.

It is an object of the present invention to provide an apparatus and method of more accurately analyzing productivity.

In order to attain the above objects of the present invention, among others, the present invention provides a method and apparatus for estimating staffing that uses a computer to estimate required staffing. The computer stores in a memory various predictive factors regarding the work presently being performed, predictive factors regarding work that has been completed and the resources allocated to that completed work that allowed for its timely completion. Using the past work predictive factors and associated allocated resources as a guide, models are created that describe the correlation between that past work and the allocated resources. Thereafter, the models are used to determine the staffing levels required to adequately perform the present work.

In a preferred embodiment, the computer stores in a memory various characteristics regarding the present patient population, such as age, sex, and physical ailment, along with a representative past patient population having similar characteristics and resources allocated to that past patient population, including staffing levels, that were required to care for the past patent population. Using the past patient characteristics and associated allocated resources as a guide, models are created that describe the correlation between that past patient population and the allocated resources. Thereafter,

the models are used to determine the staffing levels required t adequately care for the present patient population.

Furthermore, the present invention allows for the refinement of the estimation of staffing levels based upon other factors such as local practice, adjustment for changing workload patterns, desired changes in resource utilization, and new information obtained as a result of work recently completed.

By using data obtained from past and present workload, the present invention is able to make unbiased estimates of staffing requirements in order to more appropriately match the required staffing with the present work in a given facility.

By using historical patient and workload data the present invention is able to severity adjust the historical data for patient characteristics and provide an unbiased estimate of the productivity of the workforce caring for the historical patient population.

Brief Description of the Drawings

The above and other objects, features, and advantages of the invention are further described in the detailed description which follows, with reference to the drawings by way of non-limiting exemplary embodiments of the present invention, wherein like reference numerals represent similar parts of the present invention throughout several views and wherein:

Fig. 1 illustrates a computer system capable of implementing the present invention;

Fig. 2 illustrates a flowchart of the process according to the present invention of creating a model based upon a past patient population; and

Fig. 3 illustrates a flowchart of process used to predict the staffing required to support a present patient population according to the present invention.

Detailed Description of the Preferred Embodiments

Fig. 1 illustrates a staffing allocation system 100 according to the present invention. As illustrated, the system is preferably configured as a distributed computer, having separate computers 110 for data input, processing, and output. As is known, the separate computers are preferably tied together in some type of local area network, such that each computer 110 has access, through a server 120, to a central database 130 which contains the data used by the present invention in order to obtain the desired staffing allocation information. The central database 130 need not be a single physical database, but rather is properly viewed as a collection of databases that the server 120 or computers 110 have the capability of accessing in order to obtain the necessary data. Another portion of the staffing allocation system 100 according to the present invention is an application software program resident on either certain of the computers 110 or the server 120. As is known, the processor associated with one of the computers 110 or the server 120 executes program instructions that implement the features of the invention described hereinafter. It will be understo d that various

modifications to the particular system 100 can be made without departing from the intended scope of the present invention.

The staffing allocation system according to the present invention has three overall aspects. The first aspect of the staffing allocation system is the use of completed work, predictive factors relating to such completed work, and associated resource allocations for the completed work to create models that describes the relationship between the past work and the resource allocations required for completion. The second aspect of the present invention is the use of the created models to estimate the resource allocation required to complete a present work that needs to be completed, or alternatively, the work that can be completed for a given resource allocation. The third aspect of the present invention is the updating the created models based upon new or changing information. Each of these aspects will be discussed in more detail hereinafter.

Before discussing a particular embodiment of the present invention directed to the health care environment, and in particular a hospital health care environment, it should be understood that the present invention has applicability to environments that require staffing that are different than the health care environment.

In the health care environment, each patient can be viewed as having a certain amount of work that is associated with that patient. Certain characteristics of that patient are predictive of the amount of care ("work") that the given patient requires. Therefore, while the preferred embodiment discussed hereinafter will describe the present invention with regard to the staffing needed in a hospital, it will be appreciated that certain aspects of the present invention can be applied to other work environments.

A few terms that are used herein should also be understood. While the term "work" or "project" is associated with task requiring completion, such as all of the work associated with administering care to a given patient, this is distinguished from the term "workload," which is used in association with the resource allocation needed to complete the work, or, in the specific embodiment, properly administer care to the patient.

In light of the above, the preferred embodiments of the present invention will be described with reference to the work that is performed in the hospital environment and the staffing associated therewith. Initially, however, background is provided which will assist in understanding the system implemented by the preferred embodiment of the present invention.

In the hospital environment, the patient can be viewed as requiring various different staff to allocate different amounts of time (workload) to the patient in order to properly administer care to the patient. Since records are conventionally kept on each patient, the present inventors have determined it useful to view the patient as requiring a certain workload from various different staff of the hospital, such as doctors, nurses, clerical, phlebot mist, lab tech, etc. Associated with each of these different staff are workload units, or workunits, required to care for the patient. The workload units may be broken down into skill levels, such as nursing, clerical, phlebotomist, lab tech, etc. or may be overall

workload units in any unit desired as long as that workload reference data is captured in the reference database. A typical cost center breakdown for an acute care hospital is provided in Table 1 below as an indication of the type of workload data detail that is almost universally available.

TABLE 1- COST CENTER NAMES	
Intensive Care Unit	IV Therapy
Telemetry Care Unit	Respiratory Therapy
Nursing Unit A	Physical Therapy
Nursing Unit B	Cardiac Rehabilitation Therapy
Nursing Unit C	Physical Therapy
Pediatrics	Occupational Therapy
MentalCare Unit	
OB/GYN	Hand Rehabilitation Center
Level II Nursery	Pain Management
Rehabilitation Center	Rehab Home Health Center
Labor and Delivery	Communication Disorder Center
Surgery	Emergency Services
Endoscopy	ER Registration
Recovery Room	Urgent Care
Oncology Out Patient Services	MRI Scan Occupational Health
Sterile Processing	Checkup Center Program
Central Distribution	CT Scan
Clinical Lab	Cardiac Cath Lab
Front Office Clerical	Pediatric Subspecialty Clinic
Sendouts	
Phlebotomy	Well Baby Clinic
Technician	Food Services
Supervision	
Chemistry	Central Distribution – Linen
Technician	Social Services
Technologist	Purchasing
Supervision	Grounds
Microbiology	Safety/Security
Technician	Housekeeping
Technologist	Hospitality Services
Supervision	Plant Maintenance
Blood Bank	Biomedical Maintenance
Technician	General Accounting
Technologist	Patient Accounting
Supervision	Information Services
Hematology	
Technician	Admitting
Technologist	Administration
Supervision	Patient Relations
Urinalysis	Outreach Services
Technician	
Technologist	Public Relations
Supervision	Employee Assistance
Computer Services	Volunteer Services
Administration	
Pathology	Medical Library
Pulmonary Function	Medical Records

EKG EEG Radiology Ultrasound Nuclear Medicine Pharmacy	Utilization Review Transcription Medical Staff Director Nursing Administration

The cost centers, such as the clinical laboratory, may be further broken down into the level of detail that is generally available for each of the cost centers. This process is illustrated above for the clinical laboratory. Thus the workload units collected and predicted may be total workload for the patient, the clinical laboratory workload, the microbiology workload, or the microbiology technician workload, in increasing level of detail. Similar levels of detail will exist for each of the cost centers in Table I above. This level of detail is referred to as the granularity of the data. Finely granular data contains a great amount of detail and coarsely granular data has correspondingly less detail. Finely granular data can typically be obtained for the entire facility, which allows more detailed predictions.

Furthermore, associated with each patient are a variety of characteristics that are potentially predictive of the workload required of the different staff. These potentially predictive characteristics include, for example, age and sex, demographic factors such as zip code and insurance payer, operative procedures that patient is undergoing(if any), historical and diagnostic information as reflected in ICD-9-CM diagnosis codes, procedure codes, laboratory test results, and physiologic measurements for the current episode of care. In addition, if historical data is available for previous hospitalizations, ambulatory care encounters, prescriptions, laboratory test results, and so forth they will generally have predictive value for the current work requirements. The predictive characteristics may be useful for predicting the workload for either an entire episode of care or on a day-by-day or shift-by-shift basis during an episode of care.

Fig. 2 illustrates a flowchart of the process according to the present invention of creating a model based upon a past patient population. The overall process will first be described, with each of the process steps elaborated upon more fully hereinafter. Before describing the process of creating this model, it should be understood that many such created models are typically needed in order to describe an entire patient population. Thus, many different models are created, with each created model characterizing the relationship between the staffing requirements and a particular type of patient. The appropriate model is selected subsequently for prediction, based upon the target patient characteristics, from among the many models available. In step 200, a reference data set containing case data related to a past patient population is obtained so that it can be operated upon by the processor. As is known, this data set is referred to as the training set. In particular, the reference data set is typically a discharge data set, such a modifi d UB-92 discharge data set to which workload units have been linked. Other financial and clinical data elements may also be linked to this data set, as

will be discussed. The discharge data set will contain fields relating to patient demographics, and to characteristics of this episode of care. The workload units linked to the discharge data set for a particular patient may include detailed workunit descriptors or summary descriptors. In fact, there will typically be multiple detailed working descriptors, such as nursing, clerical, etc. as noted above, as well as time-segmentation of them, for example, first-day first-shift, first-day second-shift, etc. Alternatively, there may only be a single aggregate value of workunits depending upon the level of granularity available in the data and that desired in the model and in the predictions.

An exemplary discharge data set for a single discharged patient with exemplary linked workload units, is illustrated in Table II below, with this table providing the short name for an exemplary set of fields that are associated with a single patient, and a description of that field. Appendix A provides more detailed descriptions for certain of these and other exemplary fields associated with the discharge data set.

TABLE II

HOSP_ID	hospital identification
HOSP_CNTY	hospital county
DC_DATE	discharge date
PAT_ID	patient identification
DOB	date of birth
AGE_YRS	age in years
AGE_MOS	age in months
AGE_DAYS	age in days
SEX	sex
RACE	race
ZIPCODE	zipcode
COUNTY	county
PAYOR1-n	payor
ADM_DOW	admit day of week
ADM_TYPE	admit type
ADM_FROM	admit from
DC_DISPO	discharge disposition
HOSP_SRVC	Hospital service
MD_ATTEND	attending doctor
MD_PROC1	procedure doctor
LOS	length of stay
SA_LOS	severity adjusted length of stay
DRG	DRG
MDC	major diagnostic category
DX1-DX12	ICD-9-CM Diagnosis codes
PROC1-PROC8	ICD-9-CM Procedure Codes.
PRC1_DAY1	procedure One day
APGAR_5MIN	APGAR score at 5 minutes
BIRTH_WT	birth weight
DRG_ORIG	DRG assigned by healthcare org.
SEVERITY	severity score if associated with DRG
CHARGE_TOT	total charge in dollars
SA_CHARGE	severity adjusted cost

COST_TOT	total cost
SA_COST	severity adjusted total cost
COST_DIR	total direct cost
SA_COSTDIR	severity adjusted direct cost
ADM_SVRTY	severity on admission
A1-A12	diagnoses present on admission flags
AUDIT	audit status
LINKS 1-n	link to other data sets
DEATH	died during stay
MCARE_DRG	medicare assigned DRG
MCARE_RISK	medicare risk
MCAID_PERD	medicaid per diem
MCAID_RISK	medicaid risk
COMM_PERD	commerical per diem
COMM_RISK	commercial risk
UNINSURED	uninsured
NEWBORN	newborn
OB	obstetric episode of care
CARDIAC	cardiac episode of care
PSYCH	psychological episode of care
C_SECTION	c-section during episode of care
DC_HOME	discharged home
DC_HERE	discharged to other department
FROM_HERE	admitted from another department
FROM_ER	admitted from emergency room
EXP_LOS	expected length of stay
EXP_CHARGE	expected charge
EXP_COST	expected cost
HOSP_TYPE	hospital type
HOSP_SYS	hospital system
PAY_TYPE	type of payment
DRG_RDRG	R-DRG assigned by R-DRG grouper
DRG_APRDRG	APR-DRG assigned by APR-DRG grouper
SEV_RDRG	R- DRG severity
SEV_APRDRG	APR-DRG severity
ECODE1-ECODE3	Specialized trauma diagnosis codes
WORK_C1- WORK_Cn	workload units according to invention
WORK_CT	total workload according to invention
UNUSED_1- UNUSED_n	additional fields available

While certain fields, such as age and sex will typically be on all discharge data sets, the contents of the fields in the discharge data sets will differ considerably depending on the reason the patient was in the hospital, which will have caused the patient episode of care characteristics to differ from one patient to another. Other fields that may exist within a patient's discharge data set are provided in the table attached at Appendix A, with an exemplary data dictionary describing these example fields.

The discharge data set may also contain or be linked to another data set, such as a data set that contains laboratory test results such as blood sugar, white count, or hematocrit. Or a data set that contains pharmacy prescription data such as medication, dose, and time of administration. Or a data

set that contains physiologic monitoring data such as blood pressure, respiratory rate or temperature, and other data sets that contain such similar data as is collected during a hospital episode of care.

The discharge data set may also contain such information for multiple episodes of care for patients who were hospitalized more than once during the period covered by the data set and may be further linked to outpatient and ambulatory data, insurance data, and similar healthcare related data.

The entire population of patients discharged from a healthcare system or a hospital for a year or more is preferably used, such that the numbers of past patients are large enough to provide for adequate predictions of workloads for various types, or groups, of patients. Cases may be added to the reference data set from another source if necessary. For example, if a reference data set is prepared from a general hospital and the predictions are to be made in a hospital with a large neurosurgical population not reflected in the prepared reference data set, additional discharged patients from a source with a large representation of neurosurgical cases may be added to the reference data set. This allows the reference data set more closely correlate to the typical patient population in the facility for which staffing requires estimation.

In step 210, aggregations of patients that have similar characteristics are made. These groupings may be by procedure, diagnosis, or other characteristics. One grouping that is useful due to its content as well as widespread usage is that of a Diagnosis Related Group (DRG) variant; HCFA-DRG, APR-DRG, R-DRG, APG or a similar resource-based group. The processor associated with one of the computers 110 implements these groupings by using established extraction parameters that cause association of a single discharged patient and the data corresponding thereto with a particular group. It should be noted, however, that a single discharge patient can be associated with multiple groupings if the aggregating categories are different. For simplicity in the discussion hereinafter, the parameters used to establish these groupings will be called "DRG" parameters in the remainder of this discussion, but they can actually be any selection criteria. Thus, for example, with reference to the discharge data set of Table II, the DRG field is used to assign the grouping. By identifying this DRG field within the discharge data set, this allows the computer 110 to select from this discharge data set all of the cases to be associated with the DRG grouping. So, for example, all patients who are assigned to DRG 127, "Heart Failure and Shock," may be considered as a single group to develop a model for workload prediction for patients who are admitted with heart failure and shock.

Thereafter, in step 212, using the groupings that have been established, the computer 110 operates on all of the discharged patient cases, and identifies all of the appropriate groupings for those cases. This can be implemented in many ways, such as by adding group fields to the already existing data for a particular discharged patients discharge data set or extracting all patients in a group into a separate data file. No matter how implemented, there becomes established a number of different groups, each distinguished from the other by the parameters, such as the DRG parameters mentioned above.

Thereafter, in step 214, for each group, the operator or computer 110 selects a set of candidate predictive factors. A person knowledgeable about the group being considered preferably performs or monitors this selection process, or evaluates the results. While for most groups these factors will include sex and age, among others, any measured patient characteristic may be a predictive factor, including insurance payer, zip code, diagnosis, admission temperature, and others. Other than age and sex, one of the most available and powerful predictive factors, primarily due to is current widespread use as a diagnostic tool, are ICD-9-CM diagnoses. Other diagnostic categories may also be used, but standardized definitions are desirable. Any number of candidate predictive factors for a group can be used, there typically being on the order of tens or hundreds of such candidate factors.

As part of the selection of candidate predictive, step 216 follows in which certain of the candidate predictive factors may be clustered together to form a clustered candidate predictive factor. This operation is such that the existence of any one of the clustered candidate predictive factors will result in an indication of the presence of the clustered candidate predictive factor. For instance, ICD-9-CM codes 412 and V45.82 are respectively, "history of an acute myocardial infarction" and "history of a percutaneous transluminal coronary angiography." While these codes may be used as separate candidate predictive factors, they may also be clustered into a clustered candidate predictive factor, such as Cl001 for instance. The clusters may be based on clinical criteria, as here, or may be based on any other criteria of interest.

An example of clustering of candidate predictive factors for a group is shown in Table III below

TABLE III

code	title	cluster
4011	BENIGN HYPERTENSION	
4019	HYPERTENSION NOS	CL0003
402	HYPERTENSIVE HEART DIS	CL0011
4020	MAL HYPERTENSIVE HRT DIS	CL0011
40200	MAL HYPERTEN HRT DIS NOS	CL0011
40201	MAL HYPERT HRT DIS W CHF	CL0011
4021	BENIGN HYPERTEN HRT DIS	CL0011
40210	BEN HYPERTEN HRT DIS NOS	CL0011
40211	BENIGN HYP HRT DIS W CHF	CL0011
4029	HYPERTENSIVE HRT DIS NOS	CL0011
40290	HYPERTENSIVE HRT DIS NOS	CL0011
40291	HYPERTEN HEART DIS W CHF	CL0011
403	HYPERTENSIVE RENAL DIS	
4030	MAL HYPERTENS RENAL DIS	
40300	HTN MALIG RENAL DIS	
40301	MAL HTN RENAL W/ FAILURE	CL014F
4031	BENIGN HYPERT RENAL DIS	
40310	BEN HTN REN DIS W/O FAIL	
40311	BEN HTN REN DIS W/O FAIL	CL014F
4039	HYPERTENS RENAL DIS NOS	

40390	UNSPC HTN RENAL W/O FAIL	
40391	UNSPC HTN RENAL W/FAILUR	CL014F
404	HYPERTEN HEART/RENAL DIS	
4040	MAL HYPERT HRT/RENAL DIS	
40400	HTN HRT/REN W/O CHF/FAIL	
40401	HTN HEART/REN W/CHF, MAL	
40402	MAL HTN HRT/REN W/FAILUR	CL014F
40403	HTN HRT/REN W/CHF,FAILUR	CL014F
4041	BEN HYPERT HRT/RENAL DIS	
40410	BEN HTN HRT/REN W/O FAIL	
40411	BEN HTN HRT/REN W/CHF	
40412	BEN HTN HRT/REN W/FAILUR	CL014F
40413	BEN HRT/REN W/CHF,FAILUR	CL014F
4049	HYPERT HRT/RENAL DIS NOS	
40490	UNSPC HTN HRT/REN DIS	

As illustrated, clinical criteria corresponding to ICD-9-CM codes 402, 4020, 40200, 40201, 4021, 40210, 40211, 4029, 40290, and 40291 are candidate predictive factors that have been clustered together to form a clustered candidate predictive factor, labeled CL0011.

Step 218 follows thereafter, which begins the process of selecting those candidate predictive factors for the group that will be used as actual predictive factors. Since there are thousands of ICD-9-CM codes, many codes that can be clustered together, and numerous other possible patient characteristics, this can lead to there being an extremely large group of candidate diagnosis-related predictive factors. The selection of predictive factors from these candidates may be accomplished either automatically by the computer or interactively by the operator. The automatic process is similar to the interactive process that is described immediately hereafter. To initiate the process of selecting the actual predictive factors, the presence or absence of each candidate predictive factor, inclusive of the clustered candidate predictive factors, for each case in the discharge data set is summarized. Specifically, a spreadsheet or other listing is generated by the computer 110 that provides, for each candidate predictive factor, such as individual diagnoses, or collectively diagnoses that make up a cluster, such as CL0011, the number of patients who displayed that candidate predictive factor (count), the length of stay (los) and the average workload units of each type consumed by the episodes of care that had that factor associated with them (workload units). An example of the generated predictive factor summary is illustrated in Table 4 below

TABLE 4

diagnosis	title	count	los	Workload Units	cluster
41071	SUBENDO AMI/1ST EPISODE	396	6	28670	
41401	AMI, FIRST EPISODE, NOS	773	6	32796	
2720	PURE HYPERCHOLESTEROLEM	161	5	27470	

496	CHR AIRWAY OBSTRUCT NEC	78	6	31748	CL0017
41001	AMI A/L WALL/1ST EPISODE	66	6	32227	
4281	LEFT HEART FAILURE	9	10	46131	
78551	CARDIOGENIC SHOCK	52	6	45371	
42741	VENTRICULAR FIBRILLATION	42	6	40744	
5990	URIN TRACT INFECTION NOS	48	8	32272	
0414	E. COLI INFECT NOS	23	7	29506	
40291	HYPERTEN HEART DIS W CHF	17	7	28794	CL0011
4439	PERIPH VASCULAR DIS NOS	12	7	34236	CL0011
41400	AMI NOS	52	7	30591	
41041	AMI INF WALL/1ST EPISODE	284	5	33062	
4011	BENIGN HYPERTENSION	9	4	21271	
4110	POST MI SYNDROME	10	6	28500	
4254	PRIM CARDIOMYOPATHY NEC	14	6	29185	
42731	ATRIAL FIBRILLATION	151	7	39949	
4589	HYPOTENSION NOS	71	6	32617	
9981	HEMORR COMPLIC PROCEDURE	30	6	35651	
2851	AC POSTHEMORRHAG ANEMIA	67	10	63795	
4275	CARDIAC ARREST	50	5	34299	
4271	PAROX VENTRIC TACHYCARD	97	6	38100	
41402		83	7	39369	
49390	ASTHMA W/O STATUS ASTHM	12	7	39748	CL0002

The operator can then examine the list and interactively select the candidate predictive factors to examine, up to the number available. Alternatively, the candidate predictive factors can be automatically selected based upon preset criteria such as number of occurrences, weighted values such as (occurrences * workunits) to emphasize factors that have a large overall effect, or upon factors such as (average workunits-workunits)^{2}to emphasize extreme values.

Additional predictive factors may be added automatically or based upon manual selection. Some of these factors may be selected based upon the clinical status of the patient. For example, a patient who is assigned to a group of patients who had an infection may provoke selection of predictive factors to include time from presentation to antibiotic administration, use of intravenous antibiotic, presentation white blood cell count , and infectious organism identified. A patient who is admitted with a myocardial infarction my have time to thrombolytic therapy, peak blood creatine kinase MB band level (a laboratory test), and chronic coumadin medication as candidate predictive factors that are selected. And an obstetric patient may have history of prenatal care, chronic anti-epileptic drug therapy, and admission hematocrit as candidate predictive factors.

The number of predictive factors must be large enough to allow an accurate prediction, but small enough to allow generalization of the model to data other than the training data set. A rule of thumb is to start with a number of predictive factors that is no greater than one eighth the number of training

set cases. Then the number can be automatically adjusted by the implementation or manually adjusted by the operator to give an accurate and general model. This process is discussed below. Once completed, those predictive factors that are deemed most pertinent to the group being considered will be selected, and thereby become the actual predictive factors.

Step 220 follows thereafter, in which a group input table is constructed by the computer 110 based upon the actual predictive factors, the reference data set for each case in the group and the associated workload value(s) for each case. Thus, using this information, the group input table is constructed and indicates whether, for each case, the actual predictive factors exist. Each actual predictive factor is coded as a binary "1" if the patient under study displays that factor, and is coded as a binary "0" if that factor is not displayed. Continuous factors, such as age or number of previous admissions, are calculated for each case as appropriate. With respect to continuous factors, it may be determined that the full range of the factor is not necessary or desirable for prediction. For example, with respect to age, it has been determined that ages under a certain minimum, and over a certain maximum, do not always provide any further predictive value for purposes of determining staffing. Accordingly, ages over this maximum, typically 85 years, and under the minimum, such as 45 years, for general medical/surgical patients, are entered as the maximum and minimum. These age limits may vary for each group and workload predicted. An example group-input table is illustrated in Table 5, which table includes sex, trimmed age, and six other predictive factors, as well as the total workload units associated with each case in the group.

TABLE 5

eocid_num	sex	ageT	f1	f2	f3	f4	f5	f6	Workload Units
19950115 9OK9CTNKK	1	49	1	0	1	0	0	0	16867
19950115OTCTAATUT	0	69	0	0	1	0	0	0	19437
19950115-	1	76	0	0	0	0	0	0	21711
199501156OO54DOOG	0	68	0	0	1	0	0	0	13250
19950115PW9E99E4P	0	61	1	0	0	0	1	0	8311
19950115PPJRN2NVQ	0	80	1	0	1	0	0	0	15698
19950115ZZD344Z44	1	56	0	0	0	0	1	0	14534
199501157SSU78QWW	0	67	1	0	0	0	1	0	15068
19950115VFB3BVNBJ	0	71	0	0	1	0	0	1	13876
199501157SQQK7MUQ	1	61	1	1	1	0	0	0	14245
19950115W7V1SGS7K	1	44	0	1	0	0	1	0	20886
19950115EFF4EEKFF	0	66	1	1	0	0	1	0	9519
19950115JSQ50JJJQ	0	79	1	1	1	0	0	0	19724
19950115SL18X6V8X	0	47	1	0	0	0	1	0	8695
19950115ZZEXE4URR	0	70	0	0	1	0	0	0	19934
19950115E7KO8PM8Q	1	70	0	0	1	0	0	1	12908
19950115YBGHGBYHY	0	60	0	1	1	0	0	1	15938
19950115YL1WNNMM5	0	80	1	0	1	0	0	1	18784
19950115MTBATLA2M	1	80	1	0	0	0	1	0	19853
19950115GG8ESGG48	0	51	1	1	0	0	0	0	17017
19950115RURCIRUMM	0	49	1	1	0	0	0	0	13396

19950115MM22MM3ZY	0	73	1	0	0	0	0	0	18836
19950115PWQWZZ03I	1	72	1	0	0	0	1	0	17186
19950115SO2T1STFH	0	67	1	0	0	0	0	0	12437
19950115XRBCBBZZ9	0	74	1	1	1	0	0	0	9042

Step 222 follows in which the user selects the model type from a list of appropriate candidates. Types of models that exist, as is known, include a linear model type of the form:

Workload = a + b*Factor1+c*Factor2 +... +z*FactorN

Or a nonlinear model of the form:

Workload = exp-(a + b*Factor1+c*Factor2 +... +z*FactorN)

Or:

Workload = 1/(1+exp-(a + b*Factor1+c*Factor2 +... +z*FactorN))

where Factor 1, Factor2...FactorN are the binary factors, either "0" or "1," corresponding to each actual predictive factor. Workload is the actual workload required in caring for the patient and a, b, c...z are the coefficients obtained by the model as a result of operating upon the group-input table.

An exemplary result for three different linear models is provided below in Table 6, which illustrates the linear model developed for a model based on DRG 080, a model based on Procedure 8151, and a model based upon DRG 127. As illustrated, each of these has a corresponding model type, and a number of workload units will result when the data of a particular patient is applied to the model, which has the coefficients (a) through (g), in this instance, determined in the manner described above.

TABLE 6

agg_type	agg_ code	model_ type	constant (a)	term1_ code	term1_ value (b)	term2_ code
DRG	080	4	10.10755713	Sex	1.39123176	AgeT
Procedure	8151	2	809590.2806	Sex	486480.3032	AgeT
DRG	127	4	1099.4259	Sex	716.956156	AgeT

term2_ value	term3_ code	term3_ value	term4_ code	term4_ value	term5_ code	term5_ value

(c)		(d)		(e)		(f)	
0.03149705	CI001	4.96443384	CI002	-5.98450878			
53687091.2	RF211	729720.4751					
-2.1674520	41401	111.90826	CI003	-986.841154	41071	3099.89923	

term6_ code	term6_ value (g)
PrevAdm	98.9884238

Of course, other model types, such as other mathematical models or artificial neural network (neural net) model types may be selected. While a neural net model type does not lend itself to a simple numeric notation, it is implemented for prediction in the same manner as the parametric model types. The operator may choose from a number of different model types in order to select a model that provides the best fit to the relationship between the predictive factors and the actual workload data. Alternatively, this selection may be made automatically by the implementation of the invention. This is accomplished by building each of the possible model types and selecting the one that provides the best predictions, using preset criteria, such as the highest correlation coefficient for the fit of the predicted to the actual workload or the smallest total model error. Appendix B provides additional information regarding model building.

Once the model type is chosen in step 222, then step 224 follows, in which the analytical tool corresponding to the model type is used to process the group-input table in order to fit the data to the model type and build the model for the group, or to determine that the type of model selected is inappropriate and to indicate that another model should be selected. The model building essentially performs an analysis of the group data to determine if the selected model type that can be used to describe the relationships between the known actual predictive factors for past patients that make up the group, and the workload associated with that group. This model building can be performed using a variety of techniques. For the mathematical models curve fitting, methods such as a least-squares, simplex, Newton-Raphson or similar methods may be used. As is known, an objective function is defined and is minimized by these tools. In the preferred implementation, the objective function may be a function of the difference between the predicted and actual workload for all of the patients for whom the model is being developed. A common function that is used in this type f optimization is the sum of the squares of the differences between the actual and predicted workload values. The absolute value of the differences may also be used, as may other functions. If the former function is

used, the fit is the equivalent of a so-called least-squares fit of the model to the training data set. Exemplary tools that may be used to implement these fitting algorithms include custom-coded software modules, an OLE server such as Microsoft Excel with its built in analytical tools, an OLE server with add-in analytical extensions, or analytical software engines such as SPSS, SAS, Mathematica, Statistica, or other stand-alone analytical modules. The most common modules are either Excel or Excel with an analytical add in. . If the model is a neural net or other non-parametric method, step 224 invokes the neural net trainer or other non-parametric method and passes it the group-input table(Table 5 from step 220 to create the model[Correct] The training is as is known for neural nets, with one implementation of an artificial neural network with back-propagation training being equivalent to a steepest-gradient least-squares fit of the model to the training data. The model is archived for subsequent use in the database 130, being stored in the format that corresponds to the model type, such as, for example, a database format for a parametric model or in a model library format for non-parametric and neural net models.

As is known, the model may be validated against a test discharge data set to establish its generality and accuracy. In this process, the model is used to predict the workload for discharge data sets of cases of the appropriate type, for which the actual workload is known. These are called the test or validation data sets. The accuracy of the model is expressed as the aggregated difference between the predicted and the actual workloads in the validation sets, and if this value is acceptably small, the model is accepted and is used for actual predictions. If the model is inaccurate, a new model type is built or additional, fewer, or different predictive factors are selected and the same model is rebuilt until an acceptably accurate and general model is found.

It should be noted that steps 222 and 224 can be combined, such that the program searches for different model types and attempts to fit the data, such that if one model type is not acceptable, the computer 110 automatically proceeds to the next model type until an accurate and general model is developed.

Further, steps 222 and 224 may be combined to develop two or more models on a selection of patients that are initially assigned to a single group. In this instance, the group is subdivided into two or more groups based on either operator input, an automatic decision tree process performed by the implementation of the invention, or a combination of the two. If the operator selects the subdivision, this operation is equivalent to building two separate models, as described above. An exemplary situation in which the automatic decision tree process occurs is when the initial group includes both patients that have only been seen for one episode of care in the training discharge data set and also patients that have been seen multiple times. If the model building process is unable to develop an acceptable model, the group is automatically subdivided into single- and multiple- episode patients, and separate models are developed for each group. An exemplary situation in which the combined manual and decision tree process occurs is when the initial model shows a very large sex effect on the

predicted workload. In this situati n, the operator may be asked if the discharge data set should be subdivided into male and female patients, and separate models developed for each group. The new models developed would frequently have different high occurrence diagnoses or procedures for the two groups, and would have better predictive accuracy than any single model. Similar situations where multiple models are more accurate are easily envisioned. Generally, a different model is developed for each different workload unit to be predicted for a patient in the group. Thus, a model is developed for total workload, a different model is developed for nursing workload, for total laboratory workload, for microbiology workload, for microbiology technician workload, and so on, for each level of granularity and workload category of interest. Further, a different model is developed for each time scale of interest, as described below.

Once the model is obtained for each group, this model can then be used to practice another aspect of the invention, which is the actual estimation of workload based upon a present patient population. Or to practice yet another aspect of the invention, which is the severity adjustment of and productivity calculation for a historical collection of patients. The former aspect of the invention is discussed first.

After the models have been built for all groups, the staffing evaluation system 100 can use these models to predict the staffing required to support a present patient population. This process will be described with reference to Fig. 3.

Initially, patient data regarding the present patient population is obtained. This patient data corresponds to the data that was previously discussed with respect to Fig. 2. In step 310, this data is input into the prediction module of the present invention. Whereas the data set used for model building contained all possible data that was accessible and might have predictive value, the present patient data may comprise such possible data, or may be restricted to the data elements actually required to calculate the predictive factors in the accepted model.

Thereafter follows step 312, in which the computer 110 uses the patient data set and determines the appropriate group in which to list each patient. Since the groupings have already been established, it is only necessary to assign the patient to an existing group. This assignment can be automatically performed by computer 110 by assigning a group based upon a predetermined decision tree or other decision structure, which operates on the patient data set, performed manually by the operator, or, with some combination in which the computer 110 uses a decision tree to determine the most likely groups for a given patient, and the operator then selects the most appropriate group. Generally, this is an automatic process accomplished by the computer alone. With respect to the decision tree, each of the various patients are uniquely assigned to a DRG (See Table 2), so the patient will preferably be assigned to the group corresponding to that DRG. For some other groupings, such as diagnosis or procedure, a patient might fall into two or more possible groupings. In this case, the patient may be automatically assigned to a group based on a decision tree, may be assigned manually by the operator,

or may be assigned by a combination of the two proc sses. Again, the most common implementation is for the assignment to be automatic, with the assignment made to the group that has the most accurate model. This automatic assignment is used very commonly when multiple models have been automatically built by the implementation, such as sex-specific models or number-of-episode models as described above, to improve the accuracy of the predictions. For combined manual and computer operation, those groups to which the patient could be assigned can be presented to the operator, and the group to which the operator judges is most appropriate can then be selected.

Once the appropriate group is selected in step 314, then the correct model corresponding to that group is looked up. As part of this step, the factors and weights are also retrieved from the archive if the model is mathematical, or the non-parametric model is loaded in the library routine if not.

Having determined the model to use, step 316 follows, in which the patient data is used to determine the status of the actual predictive factors used for that group, and other data used by the model is obtained. Thus, for that patient, the presence or absence of diagnoses or other binary factors is used to determine the state ("0" or "1") of the binary predictive factors. Continuous predictive factors are calculated such as the age (with thresholds applied as discussed above if applicable the number of previous admissions, and similar required continuous factors are processed. And other predictive factors such as sex and zip code are processed, as required.

Thereafter follows step 318, in which the appropriate model is run to determine the estimated workload associated with that patient, for the workload category of interest. Sample patient data, actual predictive factors, and an example calculation for a linear model are illustrated in Tables 6A-6C below. Table 6A illustrates another set of linear models developed for DRG 080, Procedure 8151, and DRG 127. Table 6B illustrates a partial set of patient data for three different patients,

Table 6A

agg_type	agg_ code	model_ type	const	term1_ code	term1_ value	term2_ code
DRG	080	4	10.10755713	Sex	1.39123176	AgeT
Procedure	8151	2	809590.2806	Sex	486480.3032	AgeT
DRG	127	2	731832.2817	Sex	492991.6661	AgeT

term2_ value	term3_ code	term3_ value	term4_ code	term4_ value	term5_ code	term5_ value
0.03149705	Cl001	4.96443384	Cl002	-5.98450878		
53687091.2	RF211	729720.4751				
53687091.2	4280	563419.0014	Cl003	101047.9659	CL011	312330.0743

Table 6B

dc_date	pat_id	age_ yrs	sex	race	zipcode	payor1	los	sa_los	drg
19950215	B9GBLEE3L	69	0	1	95111	08	3	5.2	080
19950715	R0J0Q0IJV	44	1	1	95062	07	3	5.1	083
19950215	570H0U3E3	57	1	1	95110	07	15	11.6	080

mdc	dx1	dx2	dx3	dx4	dx5	dx6	dx7	dx8	dx9	dx10	dx11	dx12
04	48289	53019	412	41401	41420							
04	1363											
04	1124	3592	53019	53130	9351	41401	412	44020	V1582			

proc1	proc2	proc3
4223	3893	9915

Note that clusters Cl001 contains diagnoses 3451 and 5433; Cl002 contains diagnoses 41400, 41401, 41402,...41499. To calculate the predicted workload units for patient B9GBLEE3L in Table 6B, it is apparent that the record shows a DRG of 080. The first row of the archived model Table 6A is for DRG 080, which is applicable, and the model type is 4, which is a linear model. As shown in Table 6C, the following terms are summed in the linear model:

Table 6C

Factor	Value		Coefficient
Constant		10.108	
Sex	0	*	1.391
AgeT	69	*	0.0315
FCI001	0	*	4.9644
FCI002	1	*	-5.4895

Accordingly, the predicted workload units are 10.108 + 0 + 2.1735 + 0 – 5.4895 or 6.792 workload units.

The sequence of steps 310-318 is then repeated for eachpatient, to determine the estimated workload for a larger group, such as a ward or an entire hospital, and the workunit predictions are aggregated according to predetermined rules. This frequently amounts to simple summation, but patient care interaction factors may also be considered. Steps 310-318 are also repeated for each patient and for each different category of workload that is to be estimated and similar aggregation occurs for each different workload category. The interaction of the matrix elements of patient rows and workload category columns may demonstrate second order patient care interactions that are considered in the aggregations.

Although the workload can be estimated, there are many dynamic events that may require updating of the estimated workload, or updating of the model. Each process will now be discussed.

With respect to updating of the estimated workload, after a patient has had workload estimated, that patient's condition may change. As a result of this change, the patient's workload can be estimated again, in light of the changed condition. Accordingly, the previously estimated workload will be removed and the new estimate, based upon the new conditions, used instead.

With respect to updating of the model, all of the models have to be rebuilt periodically to reflect changing national or state (reference) medical practice and hospital work patterns. In addition, the predictions of the model may themselves be benchmarked against reference data sets to expand the goals that may be realized. Generally, this means that a model is built using very large scale data, such as national or statewide, and the model may be refined for local conditions, such as staffing levels, practice patterns, etc., by modifying the model based upon historical data for the hospital of interest, if available. This data may be long term to reflect baseline differences in the practices of the target hospital and the reference data set or may be short term data from the hospital to reflect slightly more modern practice patterns in the hospital than in the reference data set. Typically, although not necessarily, this refinement is limited to altering the predictive weights of the already determined predictive factors by no more than a preset amount, say 20% based upon the goals of the adjustment and operator knowledge of the likely magnitude of the effects. This may be implemented in the invention by repeating step 224 using a training discharge data set that contains the required data

elements but that inc rporates discharges from the hospital of interest over the time period of interest rather than the large scale discharge data. Step 224 is constrained to allow only small adjustments of the predictive weights, as is known in the field of constrained optimization. The model may be further fine-tuned using contemporaneous data (i.e. yesterday's and last week's), with the added advantage that many of the patients that will receive care in the predicted period will be represented in the refinement data set. Of course, new models for new groups can be developed as well, and groups can be consolidated and split.

If the focus of the hospital is to move towards a benchmark level of workload, then present invention can be used to estimate the desired staffing level to care for the actual mix of patients rather than the workload that is expected to be required absent any management. For such a focus, the prediction may combine benchmark data and the local data in a relationship that reflects the management focus. This may result in a one-time adjustment or in a phased adjustment period. For example 80% local data and 20% benchmark data may be used this quarter, 60% local data and 40% benchmark data next quarter, and so forth, until the benchmark data is used exclusively to predict the required workload. The latter stage is the equivalent of a finely granular targeted productivity based on objective norms. If the model used in this process is a parametric linear model, the coefficients of the model as estimated individually on the benchmark data and local data may be simply combined in the desired ratio to develop the new model. If the model is non-parametric or non-linear, the objective function of the model-building step is modified to weight the contributions of the local data and benchmark data appropriately and a new model is constructed as described above for fine-tuning an existing model.

Also, the present invention can be integrated into either a special purpose reporting tool or a more general clinical outcomes reporting tool to facilitate user predictions of workforce requirements based upon current data or to allow user analysis of workforce efficiency as compared to historical or benchmark norms. The user analysis of workforce efficiency is based upon the severity adjustment and productivity calculation features of this invention. If Steps 310 to 318 are applied to a historical population of patients rather than to a present population of patients, the expected workload that would have been required to take care of each patient, at the level of granularity of the model and the historical data, is obtained. This expected workload can be thought of as the workload to care for the patient if he/she behaved like the average patient with his/her predictive factors, and thus the expected workload has been severity adjusted for the patient's predictive factors. The productivity is calculated as the ratio of the actual workload collected with the historical data and the expected workload. This productivity can be used, as is known in business practice, to evaluate the efficiency of the workforce caring for the patients.

In the discussion above, it was also assumed for simplification of understanding that each patient was assigned to only a single group, and, therefore, a single model. This, however, is an over-

simplificati n that must be noted. In practice, there will typically need to be many different models associated with each patient. The selection of the appropriate model is based upon the level of granularity required and the accuracy of the available models. Thus, while there may be a model based upon the entire stay of a patient, there may be another model based upon the first day of care, yet another model based upon second day of care, and so forth. similarly, there can be different models based upon the type of staff, such that there is one model for nurses, another model for clerical staff, and another model for technicians. Accordingly, the models chosen will depend in large part upon the level of granularity that has been selected and the models that have been built upon that granularity. Of course, the finer the granularity, the more choices that the operator has available, or the greater detail an automatically generated report may have. If different levels of granularity exist, it is preferable to determine relationships between the different models, so that models which are not intended to be used for a given level of granularity are not used. For example, microbiology workload could be estimated in several ways, depending upon the granularity of the training data available. It could be estimated as a fixed fraction of the total laboratory workload if that were the finest granularity available, where the fraction is established by examining existing reference or facility workload data. It could be a direct model of the microbiology workload if that data were available in the training data set. Or it could be a sum of the workloads for microbiology clerical staff, technician staff, technologist staff, and all other microbiology staff. In the preferred implementation of the invention, a direct model of the desired workload category is always the preferred choice. The summation of component workload elements is the secondary choice, and the estimation of allocated workload from a less detailed model is the least desirable choice, with lower levels of detail being increasingly less desirable. Similar criteria are used to select the models for predicting workload for an entire length of stay, a single day, single shift, and so forth. The automatic selection of the appropriate model based on these criteria may be embodied in the implementation of the invention. Also, for parametric models, there will generally be a new model for each level of granularity. So a model may be required for first-day first-shift nursing, first-day first-shift clerical, and so forth. The appropriate model selected will then be based upon the input specified by the user, such that if estimates related to first-day first shift nursing are desired, then that model is chosen.

While the present invention has been described herein with reference to particular embodiments thereof, a latitude of modification, various changes and substitutions are intended in the foregoing disclosure, and it will be appreciated that in some instances some features of the invention will be employed without a corresponding use of other features without departing from the spirit and scope of the invention as set forth in the appended claims.

APPENDIX A

Data Element Name: PATIENT WORKLOAD UNITS IN CATEGORY N
Indicators Using: All patient episode of care records.
Definition: The units of work required for the patient in category n between the admission date and the discharge date.
Short Name: WORK_Cn
Format:

Length = 10.2
Type = Numeric
Occurs = 10

Allowable Values: Any number greater than or equal to zero and less than or equal to 9999999.99 with two decimal points.

Missing Data Logic: Field is Null.

***

Data Element Name: PATIENT TOTAL WORKLOAD UNITS
Indicators Using: All patient episode of care records.
Definition: The total units of work required for the patient between the admission date and the discharge date.
Short Name: WORK_CT
Format:

Length = 10.2
Type = Numeric
Occurs = 1

Allowable Values: Any number greater than or equal to zero and less than or equal to 9999999.99 with two decimal points.

Missing Data Logic: Field is Null.

***

Data Element Name:	**ADMISSION DATE**
Indicators Using:	All patient episode of care records.
Definition:	The date the patient was admitted to the health care organization for inpatient or outpatient service.
Short Name:	ADM_DATE
Format:	YYYYMMDD Length = 8 Type = Character Occurs = 1
Allowable Values:	MM = Month (01-12) DD = Day (01-31) YYYY = Year (e.g.: 1991) Valid date must be entered (e.g.: 02/30/1993 is not valid)
Missing Data Logic:	Field is considered missing if DC_DATE or LOS is missing, as it is calculated from them.

***

Data Element Name:	**DISCHARGE DATE**
Indicators Using:	All patient episode of care records.
Definition:	The month, day, year the patient was discharged from the health care organization as an inpatient or outpatient. The month and year of this data element are used to include patients into groups for comparative analysis.
Short Name:	DC_DATE
Format:	YYYYMMDD Length = 8 Type = Character Occurs = 1
Allowable Values:	MM = Month (01-12) DD = Day (01-31) YYYY = Year (e.g.: 1991) Valid date must be entered (e.g.: 02/30/93 is not valid).
Missing Data Logic:	Field is blank.

***

Data Element Name: **ICD-9-CM PROCEDURE CODES**
Indicators Using: All patient episode of care records, when applicable.
Definition: The International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), (Volume 3) is the classification system used to assign the ICD-9-CM Procedure Codes for this hospitalization.
Short Name: PROC*nn*
Format:

Length = 4
Type = Character
Occurs = 8

Allowable Values: Any valid ICD-9-CM Code.

Missing Data Logic: None Specified.

***

Data Element Name: **MEDICAL RECORD NUMBER (EOC)**
Indicators Using: All patient episode of care records.
Definition: A health care organization provided number used to identify a specific episode of care.
Short Name: PAT_ID
Format:

Length = 12
Type = Character
Occurs = 1

Allowable Values: Any.

Missing Data Logic: Field is blank.

***

Data Element Name: **PATIENT DATE OF BIRTH**
Indicators Using: All patient episode of care records.
Definition: The month, day, year the patient was born.
Short Name: DOB
Format: YYYYMMDD
Length = 8
Type = Character
Occurs = 1

Allowable Values: MM = Month (01-12)
DD = Day (01-31)
YYYY = Year (e.g.: 1991)
Valid date must be entered (e.g.: 02/30/93 is not valid).

Missing Data Logic: Field is blank.

***

Data Element Name: **PATIENT AGE IN YEARS**
Indicators Using: All patient episode of care records.
Definition: The number of years since the patient's birth, rounded down.
Short Name: AGE_YRS
Format:
Length = 3
Type = Numeric
Occurs = 1

Allowable Values: Zero to 130.

Missing Data Logic: Field is Null.

***

Data Element Name: PATIENT AGE IN MONTHS
Indicators Using: All patient episode of care records.
Definition: The number of months since the patient's birth, rounded down.
Short Name: AGE_MOS
Format:

Length = 2
Type = Numeric
Occurs = 1

Allowable Values: Zero to eleven.

Missing Data Logic: Field is Null and AGE_YRS equals zero.

***

Data Element Name: PATIENT AGE IN DAYS
Indicators Using: All patient episode of care records.
Definition: The number of days since the patient's birth, rounded down.
Short Name: AGE_DAYS
Format:

Length = 3
Type = Numeric
Occurs = 1

Allowable Values: Zero to thirty-one.

Missing Data Logic: Field is Null and AGE_YRS equals zero.

***

Data Element Name: SEX
Indicators Using: All patient episode of care records.
Definition: The sex of the patient as recorded at date of admission, outpatient services, or start of care.
Short Name: SEX
Format:

Length = 1
Type = Character
Occurs = 1

Allowable Values:

1 = Male
2 = Female
3 = Other
4 = Unknown

Missing Data Logic: Field is blank, or value does not represent distinguishable gender.

***

Data Element Name: PATIENT RACE
Indicators Using: All patient episode of care records.
Definition: The category that describes the patient's race.
Short Name: RACE
Format:

Length = 1
Type = Character
Occurs = 1

Allowable Values:

1 = White
2 = Black
3 = Native American/ Eskimo/ Aleut
4 = Asian/ Pacific Islander
5 = Other
6 = Unknown

Missing Data Logic: Field is blank, or value does not represent distinguishable race.

***

Data Element Name: **PATIENT POSTAL CODE (ZIPCODE)**
Indicators Using: All patient episode of care records.
Definition: Postal Code of the patient's primary residence, or mailing address.
Short Name: ZIPCODE
Format:

Length = 5
Type = Character
Occurs = 1

Allowable Values: Any valid U.S. Zip Code.

Missing Data Logic: Field is blank.

***

Data Element Name: **PATIENT COUNTY**
Indicators Using: All patient episode of care records.
Definition: The category that describes the patient's county of residence.
Short Name: COUNTY
Format:

Length = 5
Type = Character
Occurs = 1

Allowable Values: Refer to Dynittls.dbf for an up to date list.

Missing Data Logic: Field is blank.

***

Data Element Name: **HOSPITAL IDENTIFICATION NUMBER**
Indicators Using: All patient episode of care records.
Definition: A uniquely assigned identification number, used to reference health care organization level information. (CHW hospitals use the OSHPD Hospital ID.)
Short Name: HOSP_ID
Format:

Length = 6
Type = Character
Occurs = 1

Allowable Values: Any combination of letters and/or numbers that represents a single existing facility.

Missing Data Logic: Field is blank.

***

Data Element Name: **PAYOR**
Indicators Using: All patient episode of care records.
Definition: The category that describes the payor.
Short Name: PAYOR*nn*
Format:

Length = 2
Type = Character
Occurs = 2

Allowable Values:

01 = Medicare
02 = Medicaid
03 = Worker's Compensation
04 = County Indigent Program
05 = CHAMPUS/ CHAMPVA/ VA
06 = Other Government
07 = Commercial HMO
08 = Commercial PPO
09 = Private Insurance
10 = Blue Cross/ Blue Shield
11 = Self Pay
12 = Charity Care
13 = No Charge
14 = Other Non Governmental
15 = Medicare HMO
16 = Medicaid HMO

Missing Data Logic: Field is blank.

***

Data Element Name: **TYPE OF ADMISSION**
Indicators Using: All patient episode of care records.
Definition: The category that describes the patient's admission status.
Short Name: ADM_TYPE

Format:

Length = 1
Type = Character
Occurs = 1

Allowable Values:

1 = Scheduled at least 24 hours before surgery.
2 = Unscheduled.
3 = Infant, under 24 hours old.
4 = Unknown

Missing Data Logic: DC_DATE or ADM_DATE is not blank, and field is blank.

***

Data Element Name: **HOSPITAL SERVICE**
Indicators Using: All patient episode of care records.
Definition: Abbreviation of the service provided to the patient by the health care organization.
Short Name: HOSP_SRVC
Format:

Length = 3
Type = Character
Occurs = 1

Allowable Values: Any valid 3-digit abbreviation.

Missing Data Logic: Field is blank.

***

Data Element Name:	**ICD-9-CM DIAGNOSIS CODES**
Indicators Using:	All patient episode of care records.
Definition:	The International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), (Volume 1 and 2) is the classification system used to assign diagnosis codes for this hospitalization.
Short Name:	DX*nn*
Format:	Length = 5 Type = Character Occurs = 12
Allowable Values:	Any valid ICD-9-CM code.
Missing Data Logic:	None specified.

***

Data Element Name:	**ADMISSION DIAGNOSIS**
Indicators Using:	All patient episode of care records.
Definition:	The international Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), (Volume 1 and 2) is the classification system used to assign the admission diagnosis for this hospitalization
Short Name:	DX_ADMIT
Format:	Length = 5 Type = Character Occurs = 1
Allowable Values:	Any valid ICD-9-CM code.
Missing Data Logic:	Field is blank.

Data Element Name: **CONSULTING DOCTOR**
Indicators Using: All patient episode of care records.
Definition: Five digit health care organization assigned code which represents the doctor who consulted on this episode of care. Field is encrypted.
Short Name: MD_CONS*nn*

Format:

Length = 5
Type = Character
Occurs = 5

Allowable Values: Any valid, encrypted doctor code.

Missing Data Logic: None specified.

***

Data Element Name: **PROCEDURE DOCTOR**
Indicators Using: All patient episode of care records.
Definition: Five digit, health care organization assigned code which represents the doctor who performed the corresponding ICD-9-CM procedure during this episode of care. Field is encrypted.
Short Name: MD_PROC*nn*
Format:

Length = 5
Type = Character
Occurs = 8

Allowable Values: Any valid, encrypted doctor code.

Missing Data Logic: Field is blank, and corresponding ICD-9-CM procedure field is not blank.

Data Element Name: ATTENDING DOCTOR
Indicators Using: All patient episode of care records.
Definition: Five digit, health care organization assigned code which represents the patient's attending doctor for this episode of care. Field is encrypted.
Short Name: MD_ATTEND

Format:

Length = 5
Type = Character
Occurs = 1

Allowable Values: Any valid, encrypted doctor code.

Missing Data Logic: Field is blank.

***

Data Element Name: HCFA DIAGNOSIS RELATED GROUP
Indicators Using: All patient episode of care records.
Definition: The HCFA classification for the admission diagnosis of this hospitalization.
Short Name: DRG
Format:

Length = 3
Type = Character
Occurs = 1

Allowable Values: Any valid HCFA DRG Code. (Usually between 1 and 500.)

Missing Data Logic: Field is blank.

Data Element Name: MAJOR DIAGNOSTIC CATEGORY
Indicators Using: All patient episode of care records.
Definition: The Major Diagnostic Category assignment for this hospitalization.
Short Name: MDC
Format:

Length = 2
Type = Character
Occurs = 1

Allowable Values: Any valid MDC Code. (Usually between 1 and 25, or 99.)

Missing Data Logic: Field is blank.

***

Data Element Name: PATIENT LENGTH OF STAY
Indicators Using: All patient episode of care records.
Definition: The number of days between the admission date and the discharge date.
Short Name: LOS
Format:

Length = 4
Type = Numeric
Occurs = 1

Allowable Values: Any whole number greater than or equal to zero.

Missing Data Logic: Field is Null.

***

Data Element Name: SEVERITY ADJUSTED LENGTH OF STAY
Indicators Using: All indicators measuring risk adjusted values.
Definition: The patient's Length of Stay modified by the patient's severity score and APR-DRG assignment, as determined by the 3M APR-DRG grouper commercial software that that assigns the APR-DRG based on the other data in the database.
Short Name: SA_LOS
Format:

Length = 4
Type = Numeric
Occurs = 1

Allowable Values: Any whole number greater than or equal to zero.

Missing Data Logic: Field is considered missing if any of LOS, DRG_APRDRG, or SEVERITY is missing.

Data Element Name: DAYS IN INTENSIVE CARE UNIT
Indicators Using: All patient episode of care records.
Definition: The number of days this patient spent in the Intensive Care Unit during this episode of care.
Short Name: ICU_DAYS
Format:

Length = 3
Type = Numeric
Occurs = 1

Allowable Values: Any whole number between 0 and 999, that is less than or equal to the LOS.

Missing Data Logic: Field is Null.

***

Data Element Name: DAYS IN CRITICAL CARE UNIT
Indicators Using: All patient episode of care records.
Definition: The number of days the patient spent in the Critical Care Unit during this episode of care.
Short Name: CCU_DAYS
Format:

Length = 3
Type = Numeric
Occurs = 1

Allowable Values: Any whole number between 0 and 999, that is less than or equal to the LOS.

Missing Data Logic: Field is Null.

Data Element Name: **DAYS IN TELEMETRY CARE UNIT**
Indicators Using: All patient episode of care records.
Definition: The number of days the patient spent in the Telemetry Care Unit during this episode of care.
Short Name: TCU_DAYS
Format:

Length = 3
Type = Numeric
Occurs = 1

Allowable Values: Any whole number between 0 and 999, that is less than or equal to the LOS.

Missing Data Logic: Field is Null.

***

Data Element Name: **ADMITTED FROM**
Indicators Using: All patient episode of care records.
Definition: The category that describes the location from which the patient was admitted to this health care organization.
Short Name: ADM_FROM
Format:

Length = 2
Type = Character
Occurs = 1

Allowable Values:

1 = Home
2 = Residential Care Facility
3 = Ambulatory Surgery
4 = Long Term Care
5 = Acute Inpatient Hospital
6 = Other Inpatient Hospital
7 = Newborn
8 = Prison/Jail
9 = Other

Missing Data Logic: Field is blank, or field represents a non-determinable value.

***

Data Element Name: **DISCHARGED TO**
Indicators Using: All patient episode of care records.
Definition: The category that describes the location that the patient was discharged to when leaving this health care organization.
Short Name: DC_TO
Format:

Length = 2
Type = Character
Occurs = 1

Allowable Values: Any valid DC_DISPO code that represents a discharge to a location.

Missing Data Logic: Field is blank.

***

Data Element Name: **AUTOPSY PERFORMED**

Indicators Using: All patient episode of care records.

Definition: Describes whether an autopsy was performed during this episode of care.

Short Name: AUTOPSY

Format: Y/N/U
Length = 1
Type = Character
Occurs = 1

Allowable Values:

Y = Yes
N = No
U = Unknown

Missing Data Logic: Field is blank, and DC_DISPO indicates death of patient.

***

Data Element Name:	**TISSUE/ORGAN DONATION**
Indicators Using:	All patient episode of care records.
Definition:	Describes whether the patient donated any tissue or organ(s).
Short Name:	DONOR
Format:	Y/N/U Length = 1 Type = Character Occurs = 1
Allowable Values:	Y = Yes N = No U = Unknown
Missing Data Logic:	Field is blank.

***

Data Element Name:	**BIOPSY RESULTS**
Indicators Using:	All patient episode of care records.
Definition:	The category that describes the results of a biopsy performed during this episode of care.
Short Name:	BIOPSY
Format:	Left padded to two characters with zeroes. Length = 2 Type = Character Occurs = 1
Allowable Values:	00 = Normal Tissue 10 = Abnormal tissue, not otherwise specified 11 = Abnormal tissue, inflammation 12 = Abnormal tissue, benign neoplasm 13 = Abnormal tissue, malignant neoplasm 98 = Biopsy not performed 99 = Information not available
Missing Data Logic:	Field is blank or evaluates to "99"

***

Data Element Name: **DISCHARGE DISPOSITION**
Indicators Using: All patient episode of care records.
Definition: The code indicating patient status as of the ending service date of the period covered by this episode of care.
Short Name: DC_DISPO
Format: Left padded to two characters with zeroes.
Length = 2
Type = Character
Occurs = 1

Allowable Values:

01 = Routine (Home)
02 = Acute care within this hospital
03 = Other care within this hospital
04 = Long term care within this hospital
05 = Acute care at another hospital
06 = Other care at another hospital
07 = Long term care at another hospital
08 = Residential Care Facility
09 = Prison or Jail
10 = Against Medical Advice
11 = Died
12 = Home health service
13 = Other

Missing Data Logic: Field is blank, or field evaluates to a non-determinable value.

***

Data Element Name: **NOSOCOMIAL**

Indicators Using: Nosocomial infection rate.

Definition: Nosocomial Infection identified during admission.

Short Name: NOSOCOMIAL

Format:

Length = 2
Type = Character
Occurs = 1

Allowable Values: Y/N

Missing Data Logic: Field is blank or outside the range of allowable values.

***

Data Element Name: BIRTH WEIGHT IN GRAMS
Indicators Using: All patient episode of care records.
Definition: The patient's weight at birth, in grams.
Short Name: BIRTH_WT
Format:

Length = 4
Type = Numeric
Occurs = 1

Allowable Values: Any whole number between 0 and 9,999.

Missing Data Logic: Field is blank and is categorized as neonatal.

***

Data Element Name: APGAR SCORE

Indicators Using: All patient episode of care records.

Definition: APGAR is the Activity, Pulse, Grimace, Appearance, and Respiration sum score given to describe neonatal condition.

Short Name: APGAR_*nn*MIN

Format: 0-10
Length = 2
Type = Numeric
Occurs = 2

Allowable Values:

Sign	0 Points	1 Point	2 Points
Activity (Muscle Tone)	Absent	Arms and Legs Flexed	Active Movement
Pulse	Absent	Below 100 BPM	Above 100 BPM
Grimace (Reflex Irritability)	No Response	Grimace	Sneeze, cough, pulls away
Appearance (Skin Color)	Blue-gray, pale all over	Normal, except for extremities	Normal over entire body
Respiration	Absent	Slow, irregular	Good, crying

Missing Data Logic: Field is blank and DX_ADMIT denotes birth.

***

Data Element Name: **PATIENT COMPLAINTS**
Indicators Using: All patient episode of care records.
Definition: The category that describes a complaint the patient may have had during this hospitalization.
Short Name: COMPLAINTS
Format:

Length = 2
Type = Character
Occurs = 1

Allowable Values: Y/N.

Missing Data Logic: Field is blank or outside the range of allowable values.

***

Data Element Name: **INCIDENT**
Indicators Using: All patient episode of care records.
Definition: The category that describes any incident that may have occurred during this hospitalization.
Short Name: INCIDENT
Format:

Length = 2
Type = Character
Occurs = 1

Allowable Values: Unknown.

Missing Data Logic: None specified.

***

Data Element Name: RISK MANAGEMENT
Indicators Using: All patient episode of care records.
Definition: The category that describes the level of Risk Management applied during this episode of care.
Short Name: RISK_MAN
Format:

Length = 2
Type = Character
Occurs = 1

Allowable Values: Y/N

Missing Data Logic: Field is blank or outside the range of allowable values.

***

Data Element Name: UTILIZATION REVIEW
Indicators Using: All patient episode of care records.
Definition: The category that describes whether a utilization review took place for this episode of care.
Short Name: UTIL_REV
Format:

Length = 2
Type = Character
Occurs = 1

Allowable Values: Y/N

Missing Data Logic: Field is blank or outside the range of allowable values.

***

Data Element Name: **MED_HX**
Indicators Using: None.
Definition: Significant medical history present.
Short Name: MED_HX
Format:

Length = 2
Type = Character
Occurs = 1

Allowable Values: Y/N

Missing Data Logic: Field is blank or outside the range of allowable values.

***

Data Element Name: **QUALITY ASSURANCE DEPARTMENT REVIEW**
Indicators Using: All patient episode of care records.
Definition: The category that describes whether the episode of care was reviewed by this health care organization's quality assurance department.
Short Name: QA_REVIEW
Format:

Length = 2
Type = Character
Occurs = 1

Allowable Values: Y/N

Missing Data Logic: Field is blank or outside the range of allowable values.

***

Data Element Name: **CASE MANAGER**
Indicators Using: All patient episode of care records.
Definition: The ID number of the case manager responsible for this episode of care.
Short Name: CASE_MANGR
Format:

Length = 3
Type = Character
Occurs = 1

Allowable Values: Any 1-3 digit alphanumeric code.

Missing Data Logic: None specified.

***

Data Element Name: **CODER**
Indicators Using: All patient episode of care records.
Definition: The ID number of the initial coder of this episode of care.
Short Name: CODER
Format:

Length = 3
Type = Character
Occurs = 1

Allowable Values: Any 1-3 digit alphanumeric code.

Missing Data Logic: None specified.

***

Data Element Name: **TOTAL CHARGES TO THE PATIENT**
Indicators Using: All patient episode of care records.
Definition: The total number of dollars billed for this episode of care.
Short Name: CHARGE_TOT
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: 0-99,999,999

Missing Data Logic: Field is Null.

***

Data Element Name: **SEVERITY ADJUSTED TOTAL CHARGES**
Indicators Using: All patient episode of care records.
Definition: The total of dollars billed for this episode of care, adjusted to account for severity.
Short Name: SA_CHARGE
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: 0-99,999,999

Missing Data Logic: Field is Null.

***

Data Element Name: TOTAL COST
Indicators Using: All patient episode of care records.
Definition: The total amount of costs incurred by this health care organization in regards to this episode of care.
Short Name: COST_TOT
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: 0-99,999,999

Missing Data Logic: Field is Null.

***

Data Element Name: SEVERITY ADJUSTED TOTAL COST
Indicators Using: All patient episode of care records.
Definition: The total amount of costs incurred by this health care organization in regards to this episode of care, adjusted to account for severity.
Short Name: SA_COST
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: 0-99,999,999

Missing Data Logic: Field is Null.

***

Data Element Name: REVENUE
Indicators Using: All patient episode of care records.
Definition: The total revenue attributed to this episode of care.
Short Name: REVENUE
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: 0-99,999,999

Missing Data Logic: Field is Null.

***

Data Element Name: ANCILLARY COST BUCKETS

Indicators Using: All patient episode of care records.

Definition: Ancillary cost buckets each contain a single monetary value which represents a portion the total costs incurred by the health care organization during this episode of care. The current breakdown is as follows:

Specific Element	Current Category	Current Description
ANC_COST0	8810	Central Supplies Costs
ANC_COST1	8020	Pharmacy Costs
ANC_COST2	8040	Imaging (x-ray) Costs
ANC_COST3	8030	Laboratory Costs
ANC_COST4	8050	Therapy Costs
ANC_COST5	8060	Procedural Costs
ANC_COST6	8070	Surgical Costs
ANC_COST7	8890	Telemetry Care Unit Costs
ANC_COST8	8080	Intensive Care Unit Costs
ANC_COST9	8100	Ward Costs

Short Name: ANC_COST*nn*

Format:

Length = 6
Type = Numeric
Occurs = 10

Allowable Values: Any whole number greater or equal to zero and less than the total direct cost (COST_DIR) of this episode of care.

Missing Data Logic: Field is Null or contains data outside the valid range.

***

Data Element Name: ANCILLARY COST DESCRIPTION
Indicat rs Using: All patient episode of care records.
Definition: A four-digit code that describes the category that describes the corresponding ancillary cost bucket.
Short Name: ANC_AC*nn*
Format:

Length = 4
Type = Numeric
Occurs = 10

Allowable Values:

Category	Description
8810	Central Supplies Costs
8020	Pharmacy Costs
8040	Imaging (x-ray) Costs
8030	Laboratory Costs
8050	Therapy Costs
8060	Procedural Costs
8070	Surgical Costs
8890	Telemetry Care Unit Costs
8080	Intensive Care Unit Costs
8100	Ward Costs

Missing Data Logic: Field is Null and corresponding Ancillary Cost element (ANC_COST*nn*) is NOT Null, or, Field does not represent a determinable ancillary description category.

***

Data Element Name: TOTAL DIRECT COST
Indicators Using: All patient episode of care records.
Definition: The monetary value that describes the total direct costs incurred by this health care organization during this episode of care.
Short Name: COST_DIR
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: Any whole number that is greater than or equal to zero, and less than the Total Costs (COST_TOT) of this episode of care.

Missing Data Logic: Field is Null and COST_TOT is NOT Null, or Field is outside the range of allowable values.

***

Data Element Name: **TOTAL SALARY COSTS**
Indicators Using: All patient episode of care records.
Definition: The monetary value that describes the total salary costs incurred by this health care organization during this episode of care.
Short Name: COST_SAL
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: Any whole number greater than or equal to zero, and less than the Total Costs (COST_TOT) of this episode of care.

Missing Data Logic: Field is Null and COST_TOT is NOT Null, or Field is outside the range of allowable values.

***

Data Element Name: **TOTAL VENTILATOR DAYS**
Indicators Using: All patient episode of care records.
Definition: The total number of days the patient was on a ventilator during this episode of care.
Short Name: DAYS_VENT
Format:

Length = 2
Type = Numeric
Occurs = 1

Allowable Values: 0-99

Missing Data Logic: Field is Null and one or more of the ICD-9CM codes indicate ventilator administration.

***

Data Element Name: **CONTRACT**
Indicators Using: None.
Definition: Identification number of primary contract insurer with primary coverage for this episode of care.
Short Name: CONTRACT
Format:

Length = 5
Type = Character
Occurs = 1

Allowable Values: Any valid insurer contract ID.

Missing Data Logic: Field is blank or does not represent a valid insurer contract ID.

***

Data Element Name: **READMIT**
Indicators Using: Unplanned readmission.
Definition: Identifies this episode of care as an unplanned readmission related to a previous admission.
Short Name: READMIT
Format:

Length = 1
Type = Character
Occurs = 1

Allowable Values: Y/N

Missing Data Logic: None specified.

***

Data Element Name: **HCFA DRG SEVERITY SCORE**
Indicators Using: All patient episode of care records.
Definition: The severity score as assigned by an HCFA Grouper.
Short Name: SEVERITY
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values: 0-4

Missing Data Logic: Field is Null or outside the range of allowable values.

***

Data Element Name: **HCFA DRG SEVERITY SCORE AT ADMISSION**
Indicators Using: All patient episode of care records.
Definition: The severity score as assigned by an HCFA Grouper at time of admission.
Short Name: SEV_ADMIT
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values: 0-4

Missing Data Logic: Field is Null, or outside the range of allowable values.

Data Element Name: **SRDRG SEVERITY**
Indicators Using: None.
Definition: SRDRG Severity score.
Short Name: SEV_SRDRG
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values: Any valid SRDRG code.

Missing Data Logic: Field is blank or outside the range of allowable values.

***

Data Element Name: **RDRG SEVERITY**
Indicators Using: None.
Definition: Refined Diagnosis Related Group severity score.
Short Name: SEV_RDRG
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values: 0-3

Missing Data Logic: Field is Null or outside the range of allowable values.

***

Data Element Name: **ALL-PATIENT REFINED DRG SEVERITY SCORE**
Indicators Using: All patient episode of care records.
Definition: The severity score of this episode of care as assigned by an APR-DRG Grouper.Severity is also assigned when the case is grouped by any grouper
Short Name: SEV_APRDRG
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values: 0-4

Missing Data Logic: Field is Null or is not within the range of allowable values.

***

Data Element Name: **APACHE SEVERITY SCORE**
Indicators Using: None.
Definition: Apache ICU Severity score.
Short Name: SEV_APACHE
Format:

Length = 3
Type = Numeric
Occurs = 1

Allowable Values: 0-99

Missing Data Logic: Field is blank or outside the range of allowable values.

***

Data Element Name: CLINICAL PATHWAY
Indicators Using: All patient episode of care records.
Definition: The category that describes the clinical pathway of this episode of care. This element is User Defined.
Short Name: PATHWAY
Format:

Length = 5
Type = Character
Occurs = 1

Allowable Values: User Defined

Missing Data Logic: None specified.

***

Data Element Name: SEVERITY ADJUSTED TOTAL DIRECT COST
Indicators Using: All patient episode of care records.
Definition: The value that describes the total direct costs the health care organization incurred from this episode of care, adjusted to account for severity.
Short Name: SA_DCOST
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: Any value greater than or equal to zero that is less than the severity adjust total cost (SA_COST).

Missing Data Logic: Field is Null, or Field not within the range of allowable values.

***

Data Element Name: **ORIGINAL DIAGNOSIS RELATED GROUP**
Indicators Using: All patient episode of care records.
Definition: The Diagnosis Related Group assigned by the health care organization (rather than that assigned by the present invention) to this episode of care.
Short Name: DRG_ORIG
Format:

Length = 3
Type = Character
Occurs = 1

Allowable Values: Any valid HCFA DRG code.

Missing Data Logic: Field is blank or does not represent a valid HCFA DRG.

***

Data Element Name: **RDRG DIAGNOSIS RELATED GROUP**
Indicators Using: None.
Definition: Refined Diagnosis Related Group.
Short Name: DRG_RDRG
Format:

Length = 3
Type = Character
Occurs = 1

Allowable Values: Any valid RDRG code.

Missing Data Logic: Field is blank or outside the range of allowable values.

***

Data Element Name: **SRDRG DIAGNOSIS RELATED GROUP**
Indicators Using: None.
Definition: SRDRG Grouper assigned refined Diagnostic Related Group code.
Short Name: DRG_SRDRG
Format:

Length = 3
Type = Character
Occurs = 1

Allowable Values: Any valid SRDRG Code

Missing Data Logic: Field is blank or is outside the range of allowable values.

***

Data Element Name: **ALL-PATIENT REFINED DRG**
Indicators Using: All patient episode of care records.
Definition: The Diagnosis Related Group as assigned by an All-Patient Refined DRG Grouper.
Short Name: DRG_APRDRG
Format:

Length = 3
Type = Character
Occurs = 1

Allowable Values: Any valid APR-DRG code.

Missing Data Logic: Field is blank or does not represent a valid APR-DRG code.

***

Data Element Name: **APG**
Indicators Using: None.
Definition: The category that describes the Abulatory Patient Group assigned to this episode of care.
Short Name: APG
Format:

Length = 3
Type = Character
Occurs = 1

Allowable Values: Any valid Abulatory Patient Group code.

Missing Data Logic: Field is blank or outside the range of allowable values.

***

Data Element Name: **DIAGNOSIS PRESENT ON ADMISSION**
Indicators Using: All patient episode of care records.
Definition: A value that describes whether the corresponding ICD-9-CM Diagnosis code was present at the time of the patient's admission to this health care organization.
Short Name: *Ann*
Format: Y/N/U

Length = 2
Type = Character
Occurs = 12

Allowable Values:

Y = DX*nn* present on admission
N = DX*nn* not present on admission
U = Status of DX*nn* on admission is not known

Missing Data Logic: Field is blank or does not represent an allowable value.

***

Data Element Name: AUDIT STATUS
Indicators Using: All patient episode of care records.
Definition: The category that describes the audit status for this episode of care.
Short Name: AUDIT
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = Unaudited
1 = Audited at least once
2 = Final sign off for submission to state
9 = Unknown

Missing Data Logic: Field is Null or does not represent an allowable value.

***

Data Element Name: PATIENT DEATH (BINARY FLAG)
Indicators Using: All patient episode of care records.
Definition: The boolean description of whether the patient died during this episode of care.
Short Name: DEATH
Format:

Boolean
Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowable value.

***

Data Element Name: **MEDICARE DRG** (BINARY FLAG)
Indicators Using: None.
Definition: The Medicare-assigned Diagnosis Related Group.
Short Name: MCARE_DRG
Format:
Length = 1
Type = Numeric
Occurs = 1

Allowable Values:
0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowable value.

***

Data Element Name: **MEDICARE DRG RISK** (BINARY FLAG)
Indicators Using: None.
Definition: Boolean value that represents whether the episode of care is classified as a risk due to the Medicare Diagnosis Related Group.
Short Name: MCARE_RISK
Format:
Length = 1
Type = Numeric
Occurs = 1

Allowable Values:
0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: **MCAID_PERD** (Binary Flag)
Indicators Using: None.
Definition: Indicates that this episode of care represents a Medicare Per Diem Patient.
Short Name: MCAID_PERD
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: **MEDICAID PER DIEM RISK** (Binary Flag)
Indicators Using: None.
Definition: A boolean value that represents whether the episode of care is classified as a financial risk due to its Medicaid Per Diem contract.
Short Name: MCAID_RISK
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: COMMERCIAL PER DIEM (Binary Flag)
Indicat rs Using: None.
Definition: Indicates that the episode of care represents a Commercial Per Diem Patient.
Short Name: COMM_PERD
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: COMMERCIAL PER DIEM RISK (Binary Flag)
Indicators Using: None.
Definition: A boolean value which represents whether the episode of care classifies as a risk of financial loss due to its Commercial Per Diem contract.
Short Name: COMM_RISK
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: **PATIENT UNINSURED** (Binary Flag)
Indicators Using: All patient episode of care records.
Definition: Boolean flag describing whether the patient was uninsured during this episode of care.
Short Name: UNINSURED
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: **NEWBORN** (Binary Flag)
Indicators Using: All patient episode of care records.
Definition: Boolean flag that describes whether the patient is neonatal.
Short Name: NEWBORN
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: **EOC IS OBSTETRIC** (Binary Flag)
Indicators Using: All patient episode of care records.
Definition: Boolean flag that describes whether the episode of care is Obstetric.
Short Name: OB
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: **EOC IS CARDIOVASCULAR** (Binary Flag)
Indicators Using: All patient episode of care records.
Definition: Boolean flag that describes whether the episode of care is cardiovascular.
Short Name: CARDIAC
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: EOC IS PSYCHOLOGICAL (Binary Flag)
Indicat rs Using: All patient episode of care records.
Definition: Boolean flag that describes whether the episode of care is Psychological.
Short Name: PSYCH
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: C-SECTION (Binary Flag)
Indicators Using: All patient episode of care records.
Definition: Boolean flag that describes whether the patient underwent a C-Section.
Short Name: C_SECTION
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: **DISCHARGED HOME** (Binary Flag)
Indicators Using: All patient episode of care records.
Definition: Boolean flag that describes whether the patient was discharged to their home.
Short Name: DC_HOME
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: **DISCHARGED TO ANOTHER DEPARTMENT** (Binary Flag)
Indicators Using: All patient episode of care records.
Definition: Boolean flag that describes whether the patient was discharged to another department within this health care organization.
Short Name: DC_HERE
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: ADMITTED FROM THIS HCO (Binary Flag)
Indicators Using: All patient episode of care records.
Definition: Boolean flag that describes whether the patient was admitted from another department within this health care organization, excluding the emergency room.
Short Name: FROM_HERE
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: ADMITTED FROM EMERGENCY ROOM (Binary Flag)
Indicators Using: All patient episode of care records.
Definition: Boolean flag that describes whether the patient was admitted from the emergency room of this health care organization.
Short Name: FROM_ER
Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: **POST PROCEDURE MORTALITY IN 48 HOURS** (Binary Flag)

Indicators Using: All patient episode of care records.

Definition: Boolean flag that describes whether the patient died within 48 hours of a major surgical procedure which took place during this episode of care.

Short Name: MORT_P48

Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: **PROCEDURE COMPLICATIONS** (Binary Flag)

Indicators Using: All patient episode of care records.

Definition: Boolean flag that describes whether a complication occurred after a procedure which took place during this episode of care.

Short Name: COMPL

Format:

Length = 1
Type = Numeric
Occurs = 1

Allowable Values:

0 = False
1 = True

Missing Data Logic: Field is Null or does not represent an allowed value.

***

Data Element Name: **DIRECT VARIABLE COST**
Indicators Using: All patient episode of care records.
Definition: The monetary value that represents the sum total of all Direct Variable costs incurred during this episode of care.
Short Name: COST_DVAR
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: Any whole number that is greater than or equal to zero and less than the Total Direct Cost (COST_DIR) of this episode of care.

Missing Data Logic: Field is Null or outside the range of allowable values.

***

Data Element Name: **HOSPITAL TYPE**
Indicators Using: All patient episode of care records.
Definition: The category that describes the health care organization itself.
Short Name: HOSP_TYPE
Format:

Length = 2
Type = Character
Occurs = 1

Allowable Values:

1 = General Acute Care
2 = Skilled Nursing or Immediate Care
3 = Psychiatric Care
4 = Alcohol/Drug Rehabilitation
5 = Acute Rehabilitation Care

Missing Data Logic: Field is blank or does not represent one of the allowed values.

***

Data Element Name: **EXPECTED LOS**
Indicators Using: All patient episode of care records.
Definition: The total number of days the patient was expected to stay under the care of the health care organization during this episode of care, given the assigned Diagnosis Related Group and Severity score.
Short Name: EXP_LOS
Format:

Length = 5
Type = Numeric
Occurs = 1

Allowable Values: Any whole number that is greater than or equal to zero and less than or equal to 99,999.

Missing Data Logic: Field is Null.

***

Data Element Name: **EXPECTED CHARGES**
Indicators Using: All patient episode of care records.
Definition: The total charges expected to be issued by the health care organization for this episode of care, given the assigned Diagnosis Related Group and Severity score.
Short Name: EXP_CHARGE
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: Any whole number greater than or equal to zero and less than or equal to 99,999,999.

Missing Data Logic: Field is Null.

***

Data Element Name: **EXPECTED COST**
Indicators Using: All patient episode of care records.
Definition: The sum of all expected costs incurred by the health care organization during this episode of care, given the Diagnosis Related Group and Severity score.
Short Name: EXP_COST
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: Any whole number that is greater than or equal to zero, and less than or equal to 99,999,999.

Missing Data Logic: Field is Null.

***

Data Element Name: **EXPECTED TOTAL DIRECT COST**
Indicators Using: All patient episode of care records.
Definition: The expected sum of all Direct Costs incurred by this health care organization during this episode of care, given the assigned Diagnosis Related Group and Severity score.
Short Name: EXP_DCOST
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: Any whole number that is greater than or equal to zero, and is less than the expected Total Cost (EXP_COST).

Missing Data Logic: Field is Null or does not represent an allowable value.

***

Data Element Name: EXPECTED DIRECT VARIABLE COST

Indicators Using: All patient episode of care records.

Definition: The expected sum of all Direct Variable costs incurred by the health care organization during this episode of care, given the assigned Diagnosis Related Group and Severity score.

Short Name: EXP_DVCOST

Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: Any whole number greater than or equal to zero, and less than the expected Total Direct Cost (EXP_DCOST).

Missing Data Logic: Field is Null, or does not represent an allowable value.

***

Data Element Name: PATIENT SOCIAL SECURITY NUMBER

Indicators Using: All patient episode of care records.

Definition: The United States Government assigned social security number of the patient. This field is encrypted.

Short Name: PAT_IDS

Format:

Length = 11
Type = Character
Occurs = 1

Allowable Values: Any.

Missing Data Logic: None.

***

Data Element Name: PATIENT ENCOUNTER NUMBER
Indicators Using: All patient episode of care records.
Definiti n: The patient encounter number.
Short Name: PAT_IDX
Format:

Length = 20
Type = Character
Occurs = 1

Allowable Values: Any.

Missing Data Logic: Field is blank.

***

Data Element Name: HOSPITAL SYSTEM
Indicators Using: All patient episode of care records.
Definition: The category that describes the system to which this health care organization belongs. Usually this evaluates to the Hospital Association.
Short Name: HOSP_SYS
Format:

Length = 3
Type = Character
Occurs = 1

Allowable Values: Any valid hospital system category.

Missing Data Logic: Field is blank.

***

Data Element Name: **HOSPITAL REGION**
Indicat rs Using: All patient episode of care records.
Definition: The category that describes the hospital system assigned region in which this health care organization belongs.
Short Name: HOSP_REG
Format:

Length = 3
Type = Character
Occurs = 1

Allowable Values: Any category code that represents a valid region.

Missing Data Logic: Field is blank.

***

Data Element Name: **HOSPITAL COUNTY**
Indicators Using: All patient episode of care records.
Definition: The category or abbreviation that describes the county in which the health care organization is physically located.
Short Name: HOSP_CNTY
Format:

Length = 4
Type = Character
Occurs = 1

Allowable Values: Along with the following codes, 4-digit abbreviations are also valid:

01 = Alameda	13 = Imperial	25 = Modoc	37 = San Diego	49 = Sonoma
02 = Alpine	14 = Inyo	26 = Mono	38 = San Fransisco	50 = Stanislaus
03 = Amador	15 = Kern	27 = Monterey	39 = San Joaquin	51 = Sutter
04 = Butte	16 = Kings	28 = Napa	40 = San Louis Obispo	52 = Tehama
05 = Calaveras	17 = Lake	29 = Nevada	41 = San Mateo	53 = Trinity
06 = Colusa	18 = Lassen	30 = Orange	42 = Santa Barbara	54 = Tulare
07 = Contra Costa	19 = Los Angeles	31 = Placer	43 = Santa Clara	55 = Tuolumne
08 = Del Norte	20 = Madera	32 = Plumas	44 = Santa Cruz	56 = Ventura
09 = El Dorado	21 = Marin	33 = Riverside	45 = Shasta	57 = Yolo
10 = Fresno	22 = Mariposa	34 = Sacramento	46 = Sierra	58 = Yuba
11 = Glenn	23 = Mendocino	35 = San Benito	47 = Siskiyou	
12 = Humboldt	24 = Merced	36 = San Bernadino	48 = Solano	

Missing Data Logic: Field is blank.

***

Data Element Name: **ICD-9-CM PROCEDURE DAY**
Indicators Using: All patient episode of care records.
Definition: The number of days from the admission date (ADM_DATE) that the ICD-9-CM procedure was performed upon the patient.
Short Name: PROC*nn*_DAY
Format:

Length = 3
Type = Numeric
Occurs = 8

Allowable Values: Any whole number that is greater than zero and less than or equal to the length of stay (LOS).

Missing Data Logic: Field is Null or is outside the range of allowable values.

***

Data Element Name: **SURGICAL FLAG**

Indicators Using: All patient episode of care records.

Definition: A single character flag assigned during processing that recognizes major surgical procedures within this episode of care.

Short Name: SURGICAL

Format: P or S
Length = 3
Type = Character
Occurs = 1

Allowable Values:

P = Primary ICD-9-CM procedure is a major surgery.
S = Any one of the secondary ICD-9-CM procedures is a major surgery.

Missing Data Logic: None specified.

***

Data Element Name: **SEVERITY ADJUSTED DIRECT VARIABLE COST**
Indicators Using: All patient episode of care records.
Definition: The sum of all direct variable costs incurred by the health care organization during this episode of care, adjusted to account for the assigned Severity score.
Short Name: SA_DVCOST
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: Any whole number greater than or equal to zero, and less than the severity adjusted total direct cost (SA_DCOST).

Missing Data Logic: Field is Null or is outside the range of allowable values.

***

Data Element Name: **EXPECTED REVENUE**
Indicators Using: All patient episode of care records.
Definition: The expected revenue of this episode of care, given the assigned Diagnosis Related Group and Severity score.
Short Name: EXP_REV
Format:

Length = 8
Type = Numeric
Occurs = 1

Allowable Values: Any whole number greater than or equal to zero, and less than or equal to 99,999,999.

Missing Data Logic: Field is Null or outside the range of allowable values.

***

Data Element Name: **CAREUNIT**
Indicators Using: All patient episode of care records.
Definition: The 3-digit abbreviation of the care unit most involved in this episode of care.
Short Name: CAREUNIT
Format:

Length = 3
Type = Character
Occurs = 1

Allowable Values: Any valid ICU abbreviation.

Missing Data Logic: Field is blank or outside the range of allowable values.

***

Data Element Name: **ADMISSION DAY OF WEEK**
Indicators Using: All patient episode of care records.
Definition: The category that describes the day of the week upon which the patient was admitted.
Short Name: ADM_DOW
Format:

Length = 1
Type = Character
Occurs = 1

Allowable Values:

1 = Monday
2 = Tuesday
3 = Wednesday
4 = Thursday
5 = Friday
6 = Saturday
7 = Sunday

Missing Data Logic: Field is blank, or does not contain a valid category code.

***

Data Element Name: **EXPECTED MORTALITY**
Indicators Using: All patient episode of care records.
Definition: A value which represents the expected possibility of mortality given the Diagnosis Related Group and Severity score assigned to this episode of care.
Short Name: EXP_MORT
Format:

Length = 3
Type = Numeric
Occurs = 1

Allowable Values: Any whole number between zero and 999.

Missing Data Logic: Field is Null, or outside the range of allowable values.

***

Data Element Name: **EXPECTED COMPLICATIONS**
Indicators Using: All patient episode of care records.
Definition: A value that represents the likelihood of a complication given the Diagnosis Related Group and Severity score assigned to this episode of care.
Short Name: EXP_COMP
Format:

Length = 3
Type = Numeric
Occurs = 1

Allowable Values: Any whole number between 0 and 999.

Missing Data Logic: Field is null, or outside the range of allowable values.

***

Data Element Name: **DYNKEY**

Indicators Using: All patient episode of care records.

Definition: A Clinical Dynamics assigned record identifier that is unique to the scope of all known data.

Short Name: DYNKEY

Format:

Length = 10
Type = Numeric
Occurs = 1

Allowable Values: Any

Missing Data Logic: Field is Null.

***

Appendix B. Models and Model Building

I. General

To build a model, we start with a collection of reference data that contains outcomes and factors. Although there may be many other elements in the data, for the purpose of the building the model the data can be considered to be a table of the form

Y1	X1,1	X1,2	.....	X1,n
Y2	X2,1	X2,2	.....	X2,n
.				
.				
.				
Ym	Xm,1	Xm,2	.....	Xm,n

Here each row represents a single admission, the Yi value is the workload for the i^{th} admission, and the Xi,j is the factor value for the i^{th} admission and the j^{th} factor.

II. Mathematical Models

The general formula for describing or predicting a dependent variable, Y, that is a function of n independent variables, $X_1, X_2, \ldots X_n$, is

(1) $Y = f(X_1, X_2, \ldots X_n)$

Where f is an arbitrary function. For our purposes, Y is the workload in the desired units, and the X's are the factors that we will use to describe or predict the workload.

If the formula is of the form

(2) $Y = A + B X_1 + C X_2 + \ldots Z X_n$

The formula is said to be linear with constant coefficients. This is one of the most common models used to predict workload.

If the formula is of the form

(3) $Y = e^{(A + B X1 + C X2 + \ldots Z Xn)}$

The formula is said to be a model with transformation of the dependent variables.

If the formula is of the form

(4) $Y = 1/1 + e\text{-}^{(A + B X1 + C X2 + \ldots Z Xn)}$

The formula is said to be a logistic regression model.

These are the three most commonly used mathematical models, but the model can be any equation that relates the outcome, workload units, to the independent variables.

The model building problem is to select the coefficients A, B, etc. such that the model explains the outcome as well as possible given the known reference data. There are numerous methods to accomplish this process. These methods include multiple linear regression and curvilinear regression, and other so-called curve fitting methods. Although special cases of this problem may be solved analytically, the most general solutions use multi-step processes.

The multi-step curve fitting methods start with the definition of an objective function that is to be minimized. This objective function is the error of the model, and minimizing the objective function corresponds to minimizing the error in the model. Using the general statement of the reference data in I, and the linear model in II(2), the observed outcome is

(5) $$Yi$$

and the Modeled outcome is

(6) $$A + B\,Xi,1 + C\,Xi,2 + \ldots$$

The error can be defined in numerous ways, including the raw error

(7) $$Yi - A + B\,Xi,1 + C\,Xi,2 + \ldots,$$

The absolute error

(8) $$|Yi - A + B\,Xi,1 + C\,Xi,2 + \ldots|,$$

the squared error

(9) $$(Yi - A + B\,Xi,1 + C\,Xi,2 + \ldots)^2,$$

or the maximum error

(10) $$\max (Yi - A + B\,Xi,1 + C\,Xi,2 + \ldots).$$

The squared error is frequently used because it does not allow cancellation of positive and negative errors and because it magnifies the importance of very large errors. The model is built, then, by selecting coefficients that minimize

(11) $$\sum_{m} (Yi - A + B\,Xi,1 + C\,Xi,2 + \ldots)^2$$

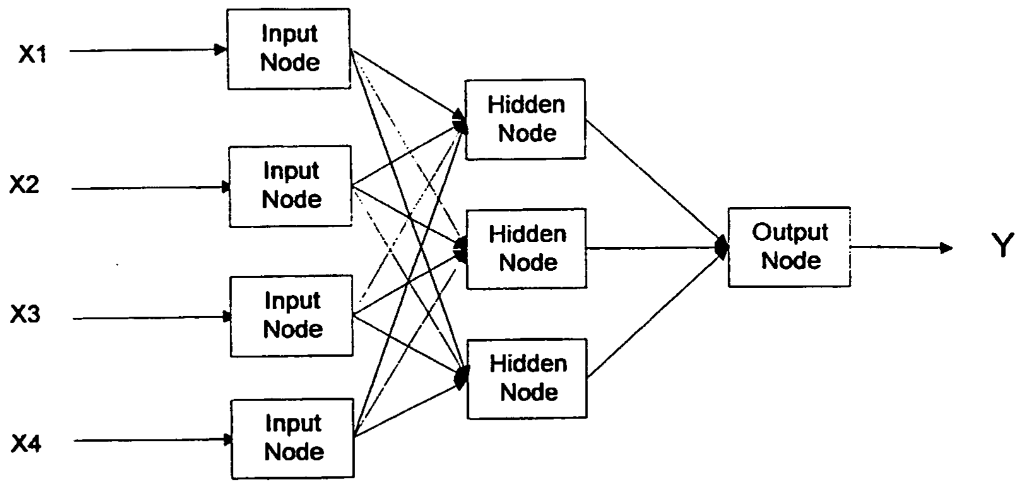
if the squared error is chosen, and in this case the model is called the least squares model.

The selection of the coefficients is actually accomplished by an initial selection more or less at random, then by varying the coefficients in a systematic way until the error in (11) reaches an acceptable (predetermined) level. There are numerous commercial software packages available to accomplish this task; WINNER uses a commercially available variant of GRG2, generalized reduced gradient method, version 2.

III. Heuristic Models

Heuristic models are those that do not use an explicit mathematical model as above, but use a trial-and-error approach to develop a model. Two of the most common heuristic models are genetic algorithms (GA) and artificial neural nets (NN).

Neural nets typically use a network model as shown below:



Neural Net

Here the inputs are the factors X1, X2, etc. and the output is Y similar to our mathematical models above. But instead of a simple linear relationship between the coefficients and factors and the output, the neural net has a weight assigned to each arrow above, i.e. to each relationship between nodes. A defined mathematical operation takes place in each of the node evaluation steps, frequently the summation of the product of each input and its corresponding weight. The output of the node is thus determined by the outputs of the nodes in the preceding layer, the weights of the relationship between each of the preceding layer nodes and the receiving node, and the operation that takes place in the node. This model allows a better prediction accuracy when there are interrelationships between the inputs as they affect the output, as might be expected with diagnoses and workload units.

The arrow weight is calculated using the training set using a specific methodology, in our case modified back propagation. The result of this training is similar to the curve fitting described above, in that the weights are varied until the error between the predicted and measured Y-values for the training set is minimized. Then the same weights are used for prediction when the inputs are fed to the NN and the output is the predicted Y.

I claim:

1. A method of estimating workload units of staffing needed at a facility to care for a person having a predetermined characteristic using a computer comprising the steps of:

obtaining a data set corresponding to a plurality of past persons having been treated for the predetermined characteristic, each of said past persons having a corresponding record that includes workload unit information and factor information relating to the predetermined characteristic;

building a model that associates the obtained workload information and the factor information in the data set;

obtaining new factor information associated with the person having the predetermined characteristic;

estimating the workload units of staffing needed to care for the person by applying the model to the new factor information associated with the person having the predetermined characteristic

2. A method according to claim 1, wherein the step of obtaining a data set includes the step of obtaining factor information relating to age and sex; and

wherein the step of building the model uses the factor information relating to age and sex.

3. A method according to claim 2 wherein the step of obtaining factor information includes obtaining factor information relating to diagnosis; and

wherein the step of building the model uses the factor information relating to diagnosis.

4. A method according to claim 2 wherein the step of obtaining factor information includes obtaining factor information relating to pharmacy prescription data and

wherein the step of building the model uses the factor information relating to pharmacy prescription data.

5. A method according to claim 2 wherein the step of obtaining factor information includes obtaining factor information relating to physiologic monitoring; and

wherein the step of building the model uses the factor information relating to physiologic monitoring.

6. A method according to claim 2 wherein the physiologic monitoring information includes one of blood pressure, respitory rate and temperature.

7. A method according to claim 2 wherein the step of obtaining factor information includes obtaining factor information relating to insurance payer; and

wherein the step of building the model uses the factor information relating to insurance payer.

8. A method according to claim 2 wherein the step of obtaining factor information includes obtaining factor information relating to clinical status; and

wherein the step of building the model uses the factor information relating to clinical status.

9. A method according to claim 1 wherein the model is one of a linear model, a nonlinear model and a neural network model.

10. A method according to claim 1 wherein the step of obtaining the new factor information include obtaining new factor information relating to age and sex.

11. A method according to claim 10 wherein the step of obtaining new factor information includes obtaining new factor information relating to past visits to the facility.

12. A method according to claim 11 wherein the new factor information relating to past visits to a facility is a number corresponding to the number of past visits to a facility.

13. A method according to claim 12 wherein the number corresponding to the number of past visits to the facility is obtained for a predetermined period of time.

14. A method according to claim 13 wherein the predetermined period of time is within the last 5 years.

15. A method according to claim 13 wherein the predetermined period of time is within the last year.

16. A method according to claim 10 wherein the step of obtaining new factor information includes obtaining new factor information relating to diagnosis.

17. A method according to claim 10 wherein the step of obtaining new factor information includes obtaining new factor information relating to pharmacy prescription data.

18. A method according to claim 10 wherein the step of obtaining new factor information includes obtaining new factor information relating to laboratory test results.

19. A method according to claim 10 wherein the step of obtaining new factor information includes obtaining new fact r information relating to physiologic monitoring.

20. A method according to claim 19 wherein the physiologic monitoring information includes one of blood pressure, respitory rate and temperature.

21. A method according to claim 10 wherein the step of obtaining new factor information includes obtaining new factor information relating to pharmacy prescription data.

22. A method according to claim 10 wherein the step of obtaining new factor information includes obtaining new factor information relating to insurance payer.

23. A method according to claim 1 wherein the predetermined characteristic is a DRG parameter.

24. A method according to claim 1, wherein the predetermined characteristic is a cluster of diagnoses.

25. A method according to claim 1 wherein the predetermined characteristic is a cluster of procedures.

26. A method according to claim 1 wherein the step of building the model includes the steps of:

identifying a plurality of candidate predictive factors using the factor information;

determining a subset plurality of predictive factors that are capable of being used to predict the association of the obtained workload information and the factor information in the data set;

selecting factor information relating to the subset plurality of predictive factors from the record of each of the past persons; and

building the model using the selected factor information relating to the subset plurality of predictive factors.

27. A method of determining a model to estimate workload units of staffing needed at a facility to care for a person having a predetermined characteristic using a computer comprising the steps of:

obtaining a data set corresponding to a plurality of past persons, each of said past persons having a corresponding record that includes workload unit information and factor information;

identifying a subset plurality of said plurality of past persons that have the predetermined characteristic as a portion of the factor information;

building a model that associates the obtained workload inf rmation and the factor information in the data set using the factor information associated with the subset plurality.

28. A method according to claim 27 wherein the step of building the model includes the steps of:

identifying a plurality of candidate predictive factors using the factor information;

determining a subset plurality of predictive factors that are capable of being used to predict the association of the obtained workload information and the factor information in the data set;

selecting factor information relating to the subset plurality of predictive factors from the record of each of the subset plurality of said plurality of past persons; and

building the model using the selected factor information relating to the subset plurality of predictive factors.

29. A method according to claim 27 further including the steps of:

obtaining a updated data set corresponding to another plurality of past persons, each of said past persons having a corresponding record that includes workload unit information and factor information;

identifying another subset plurality of said another plurality of past persons that have the predetermined characteristic as a portion of the factor information;

updating the model that associates the obtained workload information and the factor information in the data set using the factor information associated with the another subset plurality.

30. A method according to claim 29 wherein the updated data set corresponding to the another plurality of past persons comes from persons having been cared for at the facility.

31. A method according to claim 29 wherein the updated data set corresponding to the another plurality of past persons comes from persons having the predetermined characteristic.

32. A method of estimating workload units of staffing needed at a facility to care for a person having a predetermined characteristic using a computer that operates upon a model comprising the steps of:

obtaining new factor information associated with the person having the predetermined characteristic;

estimating the workload units of staffing needed to care for the person by applying the model to the new factor information associated with the person having the predetermined characteristic.

33. A method according to claim 32 wherein the step of obtaining the new factor information include obtaining new factor information relating to age and sex.

34. A method according to claim 33 wherein the step of obtaining new factor information includes obtaining new factor information relating to past visits to the facility.

35. A method according to claim 34 wherein the new factor information relating to past visits to a facility is a number corresponding to the number of past visits t a facility.

36. A method according to claim 35 wherein the number corresponding to the number of past visits to the facility is obtained for a predetermined period of time.

37. A method according to claim 36 wherein the predetermined period of time is within the last 5 years.

38. A method according to claim 36 wherein the predetermined period of time is within the last year.

39. A method according to claim 32 wherein the step of obtaining new factor information includes obtaining new factor information relating to diagnosis.

40. A method according to claim 32 wherein the step of obtaining new factor information includes obtaining new factor information relating to pharmacy prescription data.

41. A method according to claim 32 wherein the step of obtaining new factor information includes obtaining new factor information relating to laboratory test results.

42. A method according to claim 32 wherein the step of obtaining new factor information includes obtaining new factor information relating to physiologic monitoring.

43. A method according to claim 42 wherein the physiologic monitoring information includes one of blood pressure, respitory rate and temperature.

44. A method according to claim 32 wherein the step of obtaining new factor information includes obtaining new factor information relating to insurance payer.

45. A method according to claim 32 wherein the predetermined characteristic is a DRG parameter.

46. A method according to claim 32, wherein the predetermined characteristic is a cluster of diagnoses.

47. A method according to claim 32 wherein the predetermined characteristic is a cluster of procedures.

48. A method of estimating workload units of staffing needed at a facility to care for a plurality of persons using a computer comprising the steps of:

obtaining a data set corresponding to a plurality of past persons, each of said past persons having a corresponding record that includes workload unit information and factor information relating to predetermined characteristics;

determining a plurality of groups, such that each group has a plurality of common characteristics;

identifying one of the plurality of groups with which each record should be associated to obtain corresponding records for each group;

building a model for each of said groups using the obtained corresponding records for each group, each model associating the obtained workload information and the factor information in the data set;

obtaining new factor information associated with each of the plurality of persons;

determining the group with which each of the plurality of persons should be associated; and

estimating the workload units of staffing needed to care for the plurality of persons using the group that each of the persons has been determined to be associated with, the new factor information for each of the plurality of persons, and the models corresponding to the determined groups.

49. A method according to claim 41, wherein, during the step of estimating, each model operates only on the new factor information of the persons that are determined to be in the same group as that of the model.

50. A method according to claim 49, wherein, each model operates to provide partial workload information and the partial workload information is summed to obtain the workload units of staffing.

51. A method of severity adjusting workload unit information relating to a plurality of past persons using a computer comprising the steps of:

obtaining a data set corresponding to the plurality of past persons, each of said past persons having a corresponding record that includes workload unit information and factor information;

building a model that associates the obtained workload information and the factor information in the data set;

severlty adjusting the workload unit information to obtain severity adjusted workload unit information.

52. A method according to claim 51, wherein the step of obtaining a data set includes the step of obtaining factor information relating to age and sex; and

wherein the step of building the model uses the factor informati n relating to age and sex.

53. A method according to claim 51 wherein the step of obtaining factor information includes obtaining factor information relating to diagnosis; and

wherein the step of building the model uses the factor information relating to diagnosis.

54. A method according to claim 51 wherein the step of obtaining factor information includes obtaining factor information relating to pharmacy prescription data and

wherein the step of building the model uses the factor information relating to pharmacy prescription data.

55. A method according to claim 51 wherein the step of obtaining factor information includes obtaining factor information relating to physiologic monitoring; and

wherein the step of building the model uses the factor information relating to physiologic monitoring.

56. A method according to claim 51 wherein the physiologic monitoring information includes one of blood pressure, respitory rate and temperature.

57. A method according to claim 51 wherein the step of obtaining factor information includes obtaining factor information relating to insurance payer; and

wherein the step of building the model uses the factor information relating to insurance payer.

58. A method according to claim 51 wherein the step of obtaining factor information includes obtaining factor information relating to clinical status; and

wherein the step of building the model uses the factor information relating to clinical status.

59. A method according to claim 51 wherein the model is a plurality of different models, each different model associated with a corresponding different predetermined characteristic, and each of said plurality of past persons are associated with one of the corresponding different characteristics.

60. A method according to claim 59 wherein each different model is one of a linear model, a nonlinear model and a neural network model.

61. A method according to claim 60 wherein the corresponding different predetermined characteristics are one of plurality of DRG parameters.

62. A method acc rding to claim 60 wherein certain of the corresponding different predetermined characteristics include different clusters of diagnoses.

63. A method according to claim 60 wherein certain of the corresponding different predetermined characteristics include different clusters of procedures.

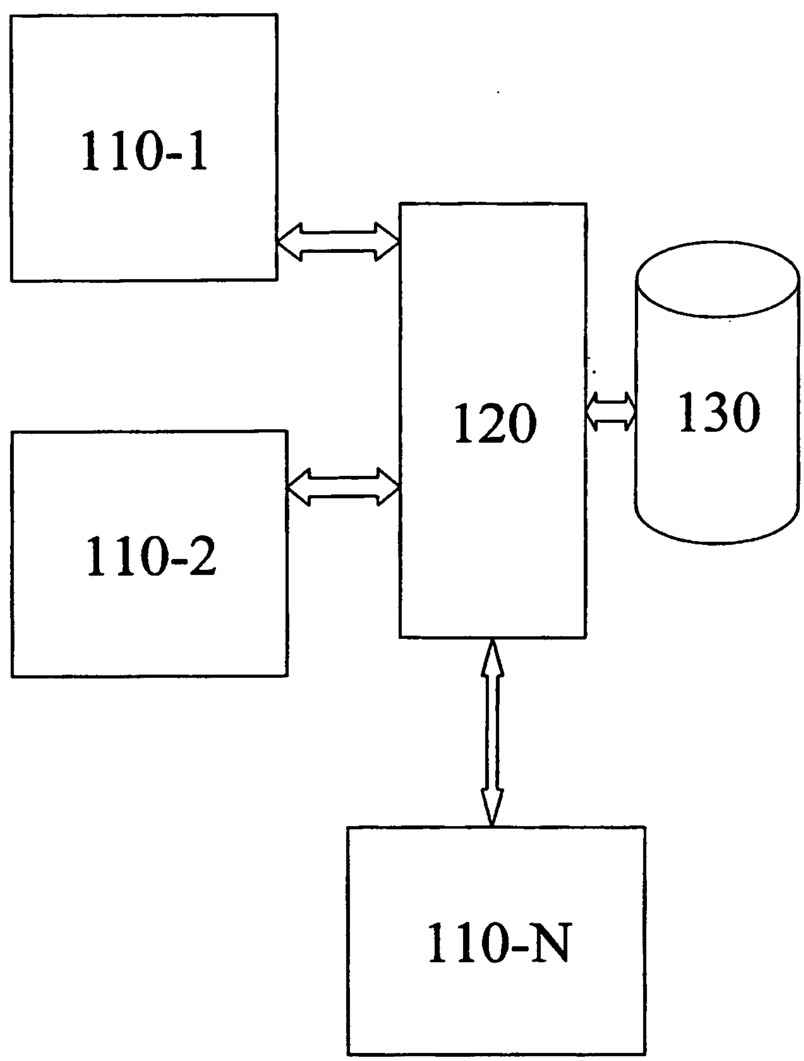


FIG. 1

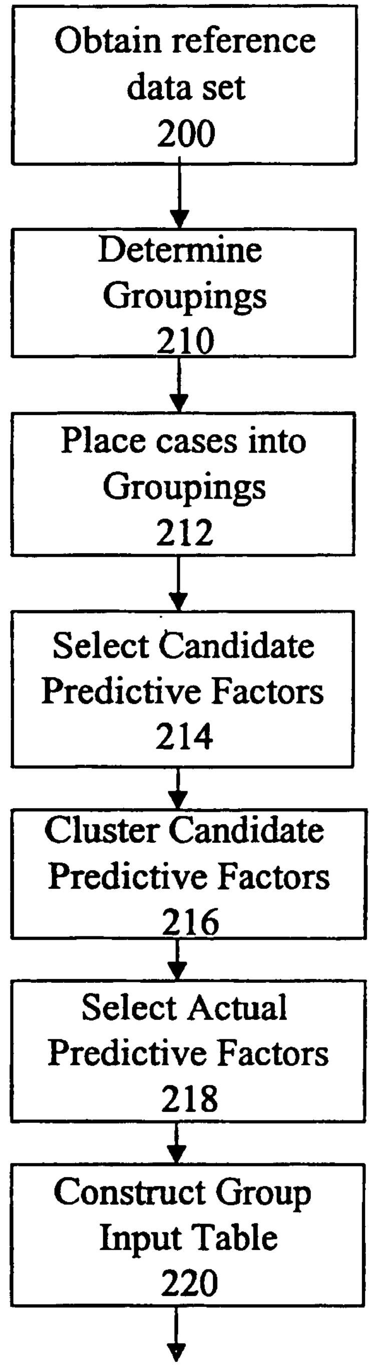

Obtain reference data set 200
Determine Groupings 210
Place cases into Groupings 212
Select Candidate Predictive Factors 214
Cluster Candidate Predictive Factors 216
Select Actual Predictive Factors 218
Construct Group Input Table 220

Fig. 2A

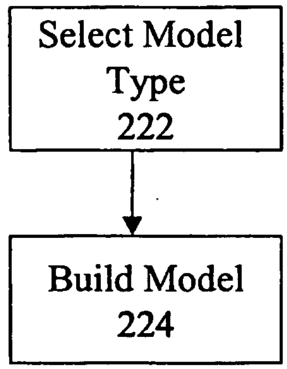


Fig. 2B

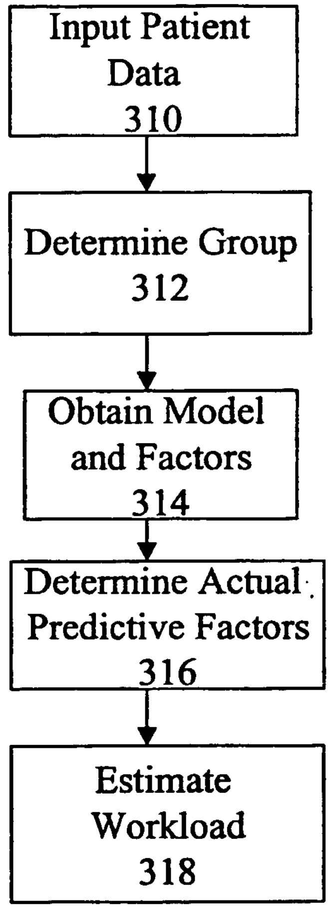

Input Patient Data 310
Determine Group 312
Obtain Model and Factors 314
Determine Actual Predictive Factors 316
Estimate Workload 318

Fig. 3